Aalto University
School of Science
Master's Degree Programme in Computational and Systems Biology (euSYSBIO)

Kul Shanker Shrestha

Germline minisatellite instability in the presence and absence of MLH1, a mismatch repair protein

Master's Thesis
Espoo, July 31, 2014

Supervisor: Prof. Juho Rousu, Aalto University
Associate prof. Arsenio Fialho, Instituto Superior Técnico
Advisor: Liisa Kauppi PhD, University of Helsinki

Aalto University
School of Science
Master's Degree Programme in Computational and Systems Biology (euSYSBIO)

ABSTRACT OF MASTER'S THESIS

Author: Kul Shanker Shrestha

Title:
Germline minisatellite instability in the presence and absence of MLH1, a mismatch repair protein

Date: July 31, 2014 **Pages:** viii + 69

Major: Computational Systems Biology **Code:** IL3013

Supervisor: Prof. Juho Rousu, Aalto University
Associate prof. Arsenio Fialho, Instituto Superior Técnico

Advisor: Liisa Kauppi PhD, University of Helsinki

MLH1 (MutL homolog 1) is a mismatch repair protein that can form heterodimer with MLH3 in meiosis. A functional MLH1-MLH3 heterodimer is essential for crossover recombination. Not much is known about the effect of loss of MLH1 function on minisatellite instability. The aim of this project is to investigate minisatellite instability in germ line cells of wild-type and $Mlh1^{-/-}$ male mice of the C57BL/6 strain, with a particular focus on minisatellite locus Prdm9 (a meiotic recombination protein in metazoans). An attempt is made to find minisatellites that function as potential DNA binding sites for PRDM9 and minisatellite instability analysis is also performed for one of them(*Ms-X165.369*). *Prdm9* locus is compared to a known stable and an unstable minisatellite, *Ms-X165.369* and two characterized microsatellites. Spermatocytes were isolated from mouse testis based on their size, and DNA was extracted. As a control, DNA extracted from somatic tissue was used. Small-pool PCR followed by Southern blotting was used to investigate and compare stability of the different minisatellite loci. *In silico* analysis was conducted to predict DNA binding motifs and DNA binding affinities for newly identified Prdm9 variants. We found germline instability at *Prdm9* locus to be higher in $Mlh1^{-/-}$ mouse than in wild type mouse , with germline instability of 16.6% and 1.4% per gamete respectively. No somatic mutation was observed at *Prdm9* locus for both the genotype. At *Ms-X165.369* locus, neither germline nor somatic mutation was observed in either of the genotype. All the PRDM9 variants were predicted to have different DNA binding motifs. DNA binding affinity of PRDM9 variants to *Ms-X165.369* was predicted to increase by at most 39% and decrease by atleast 51.5% and to the progenitor PRDM9's DNA binding motif by atleast 68%, compared to the progenitor PRDM9. Our results show that MLH1 is required for germline minisatellite stability.

Keywords: minisatellite, germline, minisatellite instability, *Prdm9*, *Ms-X165.369*, mismatch repair, MLH1, small-pool PCR, single-molecule PCR

Language: English

Acknowledgements

Foremost, I would like to thank European commission and it's Education, Audiovisual and Culture Executive Agency for providing me opportunity and scholarship to pursue my master's degree in Computational and Systems Biology (euSYSBIO). I would like to express my gratitude to Prof. Erik Aurell, Prof. Juho Rousu and Prof. Isabel Sá Correia; the coordinators of the program, for organizing an excellent program and wish all the best for its continuity. I would like to thank Karin Knutsson, Paivi Koivunen and Ana Barbosa for taking care of all the practical matters. I am also thankful to all the professors, teachers, staffs from all the three universities and everyone who supported and helped in each and every step of my learning.

I am really thankful to Prof. Juho Rousu (Aalto University) and Associate prof. Arsenio Fialho (Instituto Superior Técnico) for being my thesis supervisors.

I would like to express my gratitude to Liisa Kauppi for being my advisor and giving me opportunity to be part of her group to explore very interesting field of chromosomal instability. Her support and guidance in my summer internship and then my master's thesis was most valuable to me, without which this thesis would not have been possible at all. My thanks to all the members of Liisa's group for making my stay in the lab enjoyable and exciting. I would like to thank Maarit; most friendliest lab technician any lab can have, for managing the lab so well and teaching me those smallest details in lab practices that I would have just ignored. Juha for guidance in the lab and answering all my questions (sensible and nonsensible), and for good discussions. Imrul for his help in sample collection and fun lunch chitchats. Nanna, Miia, Barun, Elli and Hendrik for creating really comfortable working environment in the lab with friendly ambience.

At last but not the least, I would like to remember my family; without their love, care, support and blessings I would not have been where I am today.

Espoo, July 31, 2014

Kul Shanker Shrestha

Abbreviations and Acronyms

AFLPs	Amplified fragment length polymorphisms
ATP	Adenosine triphosphate
CO	Crossover
CTP	Cytidine triphosphate
DAMD-PCR	Directed amplification of minisatellite-region DNA PCR
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
DSBs	Double strand breaks
dsDNA	double strand DNA
EDTA	Ethylenediaminetetraacetic acid
ESTR	Extended simple tandem repeat
EtBr	Ethidium bromide
HCl	Hydrochloric acid
HNPCC	Nonpolyposis colorectal cancer
HR	Homologous recombination
KCl	Potassium chloride
KH2PO4	Potassium dihydro phosphate
KRAB	Kruppel associated box
MgSO4	Magnesium sulphate
MLH	MutL homolog
MMR	Miss match repair
Mms80	Mouse minisatellite 80
Ms6hm	minisatellite 6 hypermutable
MSH	MutS homologs
MSI	Microsatellite instability
Ms-X165.369	Minisatellite X165.369
MVR-PCR	Minisatellite variant repeat mapping by PCR
NaAc	Sodium acetate
NaCl	Sodium chloride

NaOH	Sodium hydroxide
NCO	Non-crossover
NHEJ	Non homologous end joining
NTP	Nucleotide tri-phosphate
PAR	Pseudoautosomal region
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Prdm9	PR domain containing 9
RAPD- PCR	Random amplified polymorphic DNA PCR
REs	Repeat elements
RFLPs	Restriction fragment length polymorphisms
rpm	Rotations per minute
RT	Room temperature
SAA	Single-strand annealing
SDS	Sodium dodecyl sulfate
SESP-PCR	Size-enrichment small-pool PCR
SM- PCR	Small molecule PCR
SNPs	Single nucleotide nolymorphisms
SP-PCR	Small pool PCR
SSC	Saline-sodium citrate
ssDNA	Single strand DNA
SVM	Support vector machine
TIM	Testes cell isolation medium
ZF	Zinc finger

Contents

Chapter 1

Introduction

1.1 Minisatellites: A subclass of repeated elements (REs)

Repeated elements, repeats of nucleotide sequences, are common feature in both prokaryotic and eukaryotic genomes. 50% of human genome and 30% of mouse genome is composed of these repeats. Repeat unit within these REs can range from one to over 100 bp and the total length of the REs can be as big as few mega-bases in mammals, but the exact number is unknown.(Jeffreys, 1997; Richard et al., 2008) DNA repeats are observed in more than one locus in genome either in tandem or in dispersed configuration.

On this basis, repeats can be divided into two large groups: tandem repeats and dispersed repeats (Richard et al., 2008)(Figure 1.1). Tandem repeats can further be grouped as gene tandems, ribosomal DNA repeat arrays, and satellite DNA. These tandem repeats are found in intergenic regions, and also in coding and non-coding regions of genes (Usdin, 2008). The class of satellite DNA consists of highly repetitive sequences, and 10% of mouse genome comprises of these repetitive sequences (Richard et al., 2008). Satellite DNA can further be subdivided into satellites, minisatellites, and microsatellites according to their repeat unit size (Richard et al., 2008). Although there is no exact demarcation in the repeat unit size for different subtypes of satellite DNA, microsatellites are generally defined as consisting of repeat units of 1 to 6 bp (Yauk et al., 2002; Bonhomme et al., 2007). Minisatellites are often GC-rich; repeat unit size can range from 10 to over 100 bp and the total can range from 100 bp to more than 20 kb (Alver et al., 2013; Bois et al., 1998; Bonhomme et al., 2007). Any repeat unit size larger than minisatellites is considered as satellites. There is also a group a satellite DNA having size

between microsatellite and minisatellite, these are known as Extended Simple Tandem Repeat (ESTR) (Yauk et al., 2002).

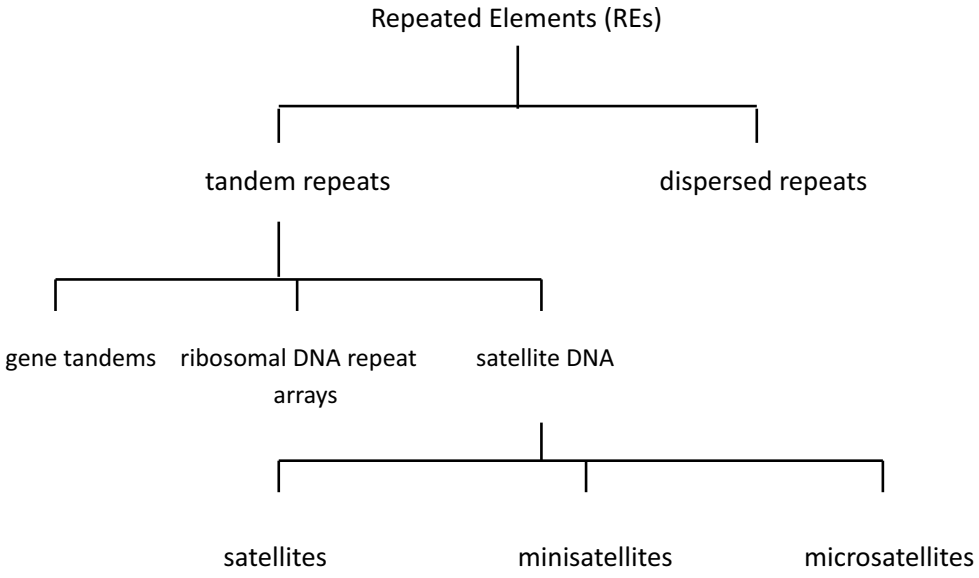

Figure 1.1: Classification of minisatellite

1.2 Minisatellite instability

Minisatellite can show variability in repeat array length or in sequence of individual repeat units, i.e. core repeat sequences. Term "Minisatellite Instability" refer to variability in repeat array, i.e. alteration (expansion/contraction) in the repeat array length. Minisatellite instability can arise from DNA replication due to slippage of DNA polymerase, malfunction or absence of Mismatch Repair (MMR) proteins or due to intra-allelic unequal crossing over (Jeffreys and Neumann, 1997). Often, if the mutated minisatellite locus is intergenic they are selectively neutral but in other cases it may lead to generation of chromosome fragility, synthesis of toxic products, synthesis of malfunctional or nonfunctional proteins in-case of intragenic minisatellite (example Insulin gene, *Prdm9*), gene silencing, modulation of transcription and translation (Usdin, 2008). These in turn can lead to several diseases such as diabetes mellitus, myoclonus epilepsy, and various types of cancer (Alver et al., 2013). In mouse germline cells , minisatellite are rather stable, with instability of less than 10^{-3} per gamete arising mainly from intra-allelic events like deletion and duplication (Bois et al., 2002; Bonhomme et al., 2007). In contrast, some human minisatellite are highly unstable mainly due to interallelic recombination events. (Bois et al., 2002)
In mouse, minisatellite locus *Ms6-hm* (minisatellite 6 hypermutable) is the most unstable locus known to date (Kelly et al., 1991, 1989) and minisatellite locus *Mms80* (Mouse Minisatellite 80) is one of the known stable mouse minisatellite (Bois et al., 2002). These minisatellite will be discussed in more detail in next section.
Through out this document, progenitor allele refers to an allele that is seen in highest frequency which is a non-mutant(wild type) form and allele with instability is termed as mutant allele.

1.3 Minisatellite loci examined in this study

We started our study with four mouse minisatellite loci: *Prdm9* (PR domain containing 9), *Ms-X165.369* (Minisatellite X165.369), *Ms6hm* (minisatellite 6 hypermutable) and *Mms80* (Mouse Minisatellite 80). The latter two has been previously characterized as the most unstable and stable minisatellite loci respectively, known to date. *Prdm9* is a meiotic recombination protein in metazoans (Oliver et al., 2009a; Segurel et al., 2011). Its germline instability in absence of functional MLH1 protein is presently not known. And, *Ms-X165.369* is a novel minisatellite locus that we discovered in this project. Due to the limited time-frame of the project and majority of time being

used for optimization of different steps, we could perform complete analyses only for *Prdm9* and *Ms-X165.369*. A Graphical representation of location of minisatellites in the chromosomes can be viewed below in Figure 1.2.

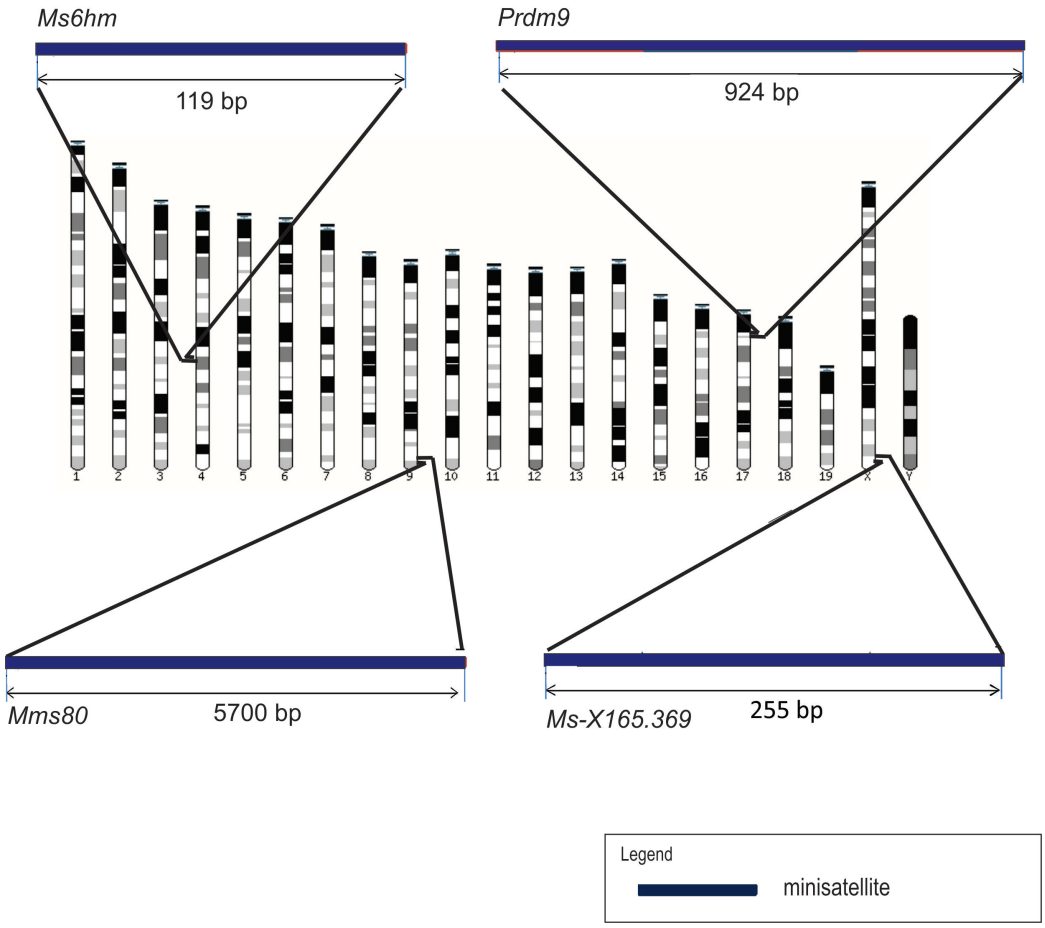

Figure 1.2: Graphical representation of the chromosomes along with the size and location of minisatellites under study. Karyotype of mouse from Flicek et al.,2014

1. *Prdm9*: PRDM9 protein is a meiotic recombination protein in metazoans which is expressed only in germline cells (Oliver et al., 2009a; Segurel et al., 2011). It has three domain: SET domain, Kruppel associated box (KRAB) domain and zinc finger (ZF)domain. The SET domain has histone H3K4 trimethyltransferase activity, the function of the KRAB domain is unknown, and the zinc finger domain is the DNA binding domain.(Hochwagen and Marais, 2010; Baudat et al., 2010)
Of particular interest in this project is the zinc finger domain of PRDM9. The zinc finger domain has a minisatellite structure consisting of a tandem array of eleven 84-bp repeats. In humans, over 20 variants of the PRDM9 ZF have been reported. Meiotic recombination events are clustered into hotspots. These hotspot locations are specified by PRDM9 by binding to specific sequences. This, in turn, could have a profound influence on germline DNA diversity. These specific sequences, i.e. DNA motifs are present in the 40% of the hotspots and have two variants in human: 13-mer motif CCnCCnTnnCGC and 17-mer motif CCCCaGTGAGCGTtgCc (Segurel et al., 2011; Hochwagen and Marais, 2010).13-mer motif is the most common of the two. *Prdm9*-null mice have been reported to show cell cycle arrest in meiotic prophase I and also impaired DSB repair (Segurel et al., 2011; Hochwagen and Marais, 2010). Alterations in the PRDM9 zinc finger (minisatellite) structure leads to changes in the protein's DNA binding affinity, which alters recombination hotspot distribution. So far, there is no direct evidence on the instability of *Prdm9* in mice, although inter strain comparisons indicate that the gene is rapidly evolving (Oliver et al., 2009b). In human sperm, *Prdm9* appears to be quite stable, despite the fact that variation in the human population suggests rapid evolution, similar to mice (Berg et al., 2010; Ponting, 2011).
In mouse, *Prdm9* gene is located in chromosome 17. *Prdm9* zinc finger minisatellite is the only intragenic minisatellite included in this study, and it is in the exon 12 of the *Prdm9* gene (Parvanov et al., 2010).

2. *Ms-X165.369* : Minisatellite *Ms-X165.369* is a novel minisatellite that we discovered during this project. Details of this minisatellite will be discussed in material and methods, and then result section.

3. *Ms6hm* (minisatellite 6 hypermutable): *Ms6hm* is located on chromosome 4 of mouse (Kelly et al., 1989). The *Ms6hm* locus contains tandem pentameric repeats of (GGGCA)n. Some classify this locus as an ESTR (Yauk et al., 2002), some as a microsatellite (Weitzmann et al., 1998) and others as minisatellite (Kelly et al., 1991). Main feature of a minisatellite is G-richness of the repeat unit, thus *Ms6hm* falls into minisatellite cate-

gory (Kelly et al., 1989). It is the most unstable minisatellite locus known to date in mouse, with a germline mutation rate of 2.5% per gamete, but the reason for instability is unknown (Kelly et al., 1991, 1989). Also, somatic mutation events have been reported in this locus (Kelly et al., 1989). This minisatellite locus has tendency to form a hairpin and two unusual intrastrand tetraplexes (Weitzmann et al., 1998). Comparisons between mouse strains of this minisatellite have shown alleles ranging from less than 200-500 repeat units (AKR mice) to more than 1000 repeat units (C57BL/6J mice) (Kelly et al., 1989).

4. *Mms80*: In mouse, the minisatellite locus *Mms80* is located in subtelomeric region of chromosome 9 (Bonhomme et al., 2007). Each repeat unit of this minisatellite is 40 nucleotides long. *Mms80* is one of the known stable mouse minisatellites. In mouse sperm, mutation rate at this locus is very low, at or below 5×10^{-6} per sperm (Bois et al., 2002). Taking into account the core repeat sequence of *Mms80* locus, it is one of the minisatellite having most variation within the core repeat sequence (Bois et al., 2002). It is highly polymorphic considering intra-repeat unit sequences.

1.4 Various modifications of PCR for minisatellite analysis

Minisatellite instability can be studied by various methods: pedigree analysis, Restriction fragment length polymorphism (RFLP) analysis or by modified PCR techniques such as Single molecule PCR (SM-PCR), Small pool-PCR (SP-PCR), Amplified fragment length polymorphisms (AFLPs), Direct amplification of minisatellite-region (DAMD-PCR) and Minisatellite variant repeat mapping by PCR (MVR-PCR) (Polyzos et al., 2006; Jeffreys et al., 1987; Monckton et al., 1993; Yauk et al., 2002; Mueller and Wolfenbarger, 1999; Jeffreys and Neumann, 1997; Heath et al., 1993). Compared to PCR based methods, pedigree analysis are time consuming and expensive as large number of individuals (F1 mice or humans) are required for a single experiment. Furthermore, PCR based techniques can also be used for analysis of somatic mutations in *in-vivo* and *in-vitro* systems. (Yauk et al., 2002) As such, modified PCR techniques are more in use at present for minisatellite instability studies.
Among the PCR-based methods, variations between the core repeat sequences in a minisatellite can be analyzed using MVR-PCR (Coleman et al., 2001; Jeffreys and Neumann, 1997; Monckton et al., 1993). Minisatellite in-

stability involving gain or loss of repeat units can be studies by SM-PCR and SP-PCR. AFLP analysis is performed for minisatellites with smaller alleles (around 1000 bp) (Mueller and Wolfenbarger, 1999). DAMD-PCR is used for intra-species minisatellite polymorphism studies, in which VNTR core sequences are used as PCR primers. DAMD-PCR is based on fact that Intra-species VNTR core sequences are similar and that between species, they are different. (Yauk et al., 2002; Heath et al., 1993)
In SM-PCR, genomic DNA is diluted to the point that each PCR has approximately one template molecule. In SP-PCR, a single reaction has a small pool of genomic DNA ranging from five to ten genomic DNA molecule per reaction. A single diploid genome of mouse is approximately 6 pg and dilutions should be made accordingly (Polyzos et al., 2006). Results by Yauk et. al., 2002 suggests that in case of high amount of DNA template molecules, the predominant allele will have dominance in amplification over the rare alleles, as such rare alleles will not amplify at all or if they amplify, it is not in the detectable range or dominant allele masks the rare allele during visualization.

1.5 Applications of minisatellites

1. Forensic investigation and genetic diversity (Population) studies : DNA fingerprinting was the first application of minisatellites(Jeffreys et al., 1987). This method was massively used in forensic investigation and in genetic diversity studies as well as in authentication of cell lines (Latif et al., 2011; Nybom et al., 2014; Jeffreys et al., 1986; Heath et al., 1993). Each individual has a unique pattern of Hypervariable minisatellites (DNA fingerprints) which can be used as identification code. After introduction of PCR based methods, DNA fingerprinting has been replaced by various PCR-based techniques such as MVR-PCR(Hopkins et al., 1994; Jeffreys et al., 1991), AFLPs (Mueller and Wolfenbarger, 1999) are used for this purpose (Nybom et al., 2014).

2. Minisatellite mapping : For uncharacterized species, minisatellite mapping can be performed as an initial step for developing whole genome sequence (Julier et al., 1990). This can be done by performing cross-hybridization using probes from other known species (Julier et al., 1990; Kelly et al., 1989).Various methods as described above can be used for this purpose.

3. Mutagenesis reporter: Rate of mutation at minisatellite loci are much greater than in non-repetitive sequences, so these loci are used as biomark-

ers, reporters and monitoring systems for various mutagenesis (both induced and spontaneous) and anticancer drug studies (Glen et al., 2008; Yauk et al., 2002; Polyzos et al., 2006; Dubrova et al., 1998; Yauk, 1998).

1.6 Introduction to Meiosis, Spermatogenesis and anatomy of mouse testis

1.6.1 Meiosis and Spermatogenesis

The meiosis cell divisions are crucial in sexual reproduction. A single 2n germline cell undergoes meiosis to produce four haploid (n) gametes (eggs and sperm). Meiosis can be divided into two phases: meiosis I and meiosis II (Figure 1.3). In meiosis I, crossing-over occurs between homologous chromosomes, consequently bringing variation due to exchange of genetic materials. Meiosis II is analogous to mitosis, in which each chromosome is segregated into sister chromatids and passed on to individual gametes.
Meiosis I can be subdivided into four stages: Prophase I, Metaphase I, Anaphase I, and Telophase I. Prophase I can be further subdivided into leptotene, zygotene, pachytene, diplotene and diakinesis. At leptotene stage chromosomes condense and form denser structure. Also developmentally programmed double strand break (DSB)formation takes place. This is followed by zygotene stage, at which stable pairing and recombination between homologous chromosomes starts to repair DSBs. At pachytene stage, homologous recombination is completed and crossovers (COs) between homologous chromosomes are matured. COs facilitates exchange of genetic material between homologous chromosomes and, in turn, reshuffles genetic variation. During diakinesis, homologous chromosomes begin to migrate apart from each other, only being held together by chiasmata (region of DNA where two non-sister homologous chromatids are physically attached and where COs occurred). In diakinesis the non-sister homologous chromatids further migrate apart, also the nucleolus completely disappears and chromosomes move towards the equatorial region of the cell. Prophase I is followed by metaphase I, in this phase chromosomes align at the metaphase plate and spindle fibers attach to their centromeres. During Anaphase I, the homologous chromosomes are completely detached from each other and each homolog moves towards opposite poles. This movement is assisted by spindle fibres. During, Telophase I, the homologous chromosomes reach the poles.

Meiosis I is followed by meiosis II, meiosis II also has four phases: Prophase II, Metaphase II, Anaphase II and Telophase II. Meiosis II cell division process is analogous to the mitotic cell division. In Prophase II, chromosomes condense and form a denser structure, spindle fibers re-appear and chromosomes start to migrate towards the equatorial plane of the cell. In metaphase II, as in metaphase I, chromosomes align in the metaphase plate and spindle fibers attach to centromeres of the chromosomes. In Anaphase II, sister chromatids detach from each other and move towards opposite poles. During Telophase II, chromatids reach the poles, the spindle fibres disappear, the nucleolus reappears along with the nuclear membrane and the chromatin de-condenses. Meiosis is followed by cytokinesis (i.e. division of cytoplasm). In the end of Meiosis, a single precursor germline cell has given rise to four haploid gametes.(Alberts et al., 2002)

Spermatogenesis is the process by which spermatogonial cells differentiate into haploid spermatozoa (male gametes) by the process of mitosis and meiosis cell division. First, spermatogonial cells undergo mitosis to produce preleptotene spermatocytes (primary spermatocytes) by the process of spermatocytogenesis. Primary spermatocytes undergo meiosis I cell division giving two secondary spermatocytes. These cells undergo meiosis II division and each cell gives rise to two spermatids by the process of spermatidogenesis. These spermatids later get mature to form motile sperm cells, by the process of spermiogenesis. (Hess and de Franca, 2008; Scudamore, 2014)

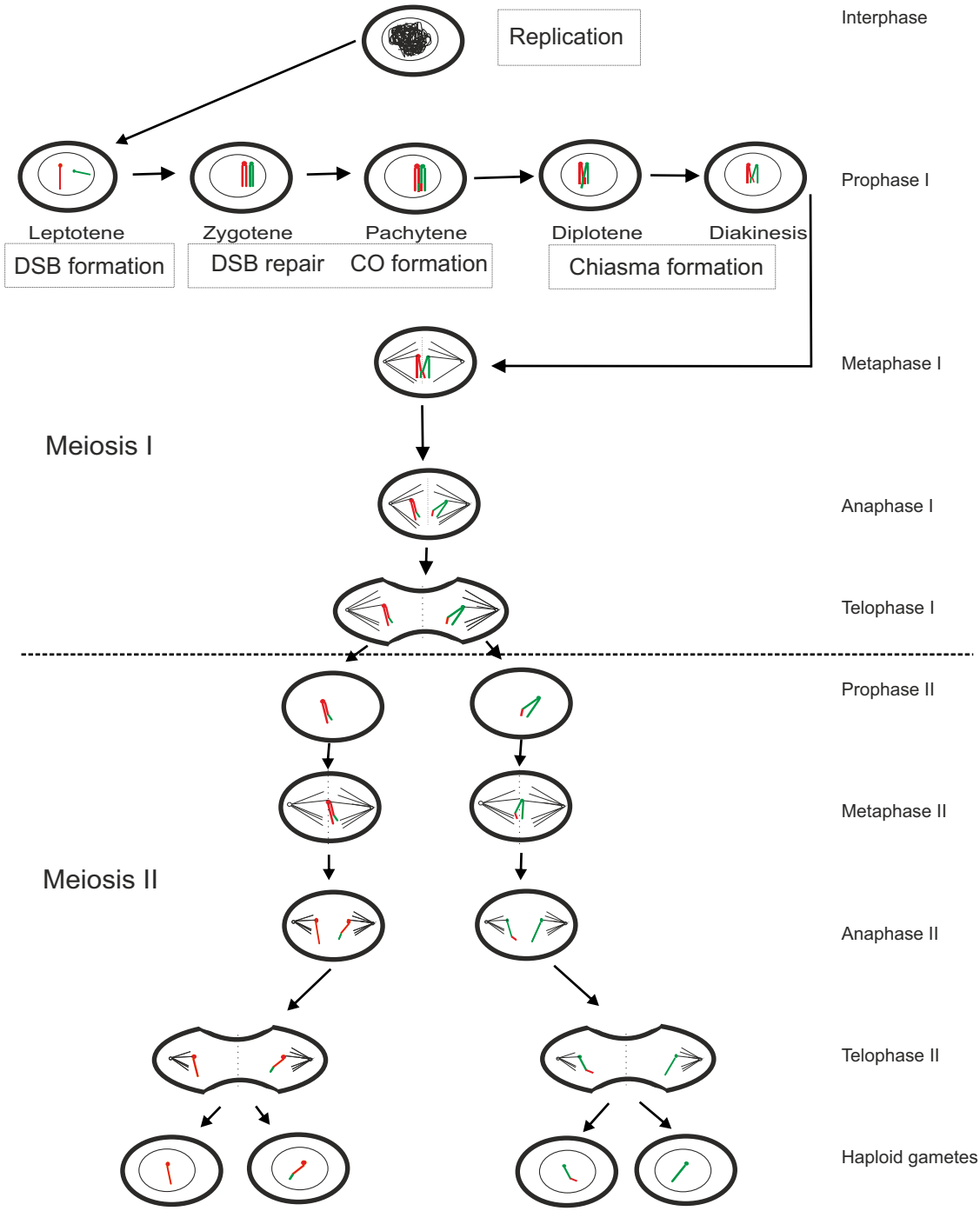

Figure 1.3: The meiosis cell divisions

1.6.2 Anatomy of mouse testis

Mice, like other male mammals, have a pair of testes. Testes are present intra-abdominally attached to adipose tissue (fat pads), on either side of the urinary bladder. Testes are present inside fibrous capsules, the tunica albuginea. Inside this capsule are highly convoluted seminiferous tubules and interstitial Leydig cells. Development of spermatogonia to mature sperm is centripetal in direction, such that various developmental stages of sperm can be observed in a cross section of a tubule (Figure 1.4). The walls of the tubules consist of a multilayered germline epithelium. The outermost layer of this epithelium is the basal lamina. In the basal lamina reside spermatogonial cells, the undifferentiated male germline cells. Inward of the spermatogonia are primary spermatocytes, followed by secondary spermatocytes. Advancing towards the lumen, spermatids are observed. The mature spermatids reside in the lumen of the tubules. The germline epithelium also has randomly scattered Sertoli cells. These cells nourish the dividing spermatocytes and spermatids. Mature sperm heads embedded into Sertoli cells can also be observed. (Hess and de Franca, 2008; Scudamore, 2014)
According to the observation made by Mays-Hoopes et. al., the average diameter of spermatogonia, spermatocytes, spermatids and Sertoli cells is $7.7\mu m$, $10.5\mu m$, $6\mu m$ and $15\mu m$ respectively (Mays-Hoopes et al., 1995) .These values are averages of 10 to 20 cell subtypes. The authors also estimated the frequency of individual cell populations for different spermatogenic lineage cells. 2.9 % of total spermatogenic lineage cells were spermatogonial cells, 22.6 % were spermatocytes at different stages and 74.4 % were spermatids at various stages (Mays-Hoopes et al., 1995)

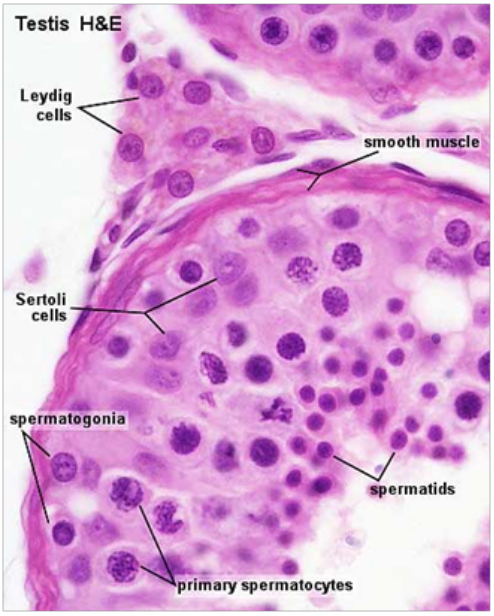

Figure 1.4: Cross section of mouse seminiferous tubule showing various developmental stage of spermatogenesis (Alver et al., 2013)

1.7 Cellular mechanisms of genome maintenance

DNA is in constant need of maintenance due to various errors (spontaneous or induced mutations)such as insertions, deletions, formation of lesions, SNPs or error in DNA replication. Therefore constant surveillance and maintenance mechanisms are required for its maintenance and integrity. These cellular mechanisms includes proofreading function of DNA polymerase, DNA mismatch repair system and homologous recombination.

1.7.1 DNA mismatch repair system

DNA mismatch repair (MMR) system in eukaryotes involves several mismatch repair proteins: MutS homologs (MSH2, MSH4,MSH5 , MSH6), MutL homologs(MLH1, MLH3, PMS1, and PMS2) (Bacher et al., 2005; Yao et al., 1999), MutH homolog. In bacteria, MutS is responsible for mismatch recognition. MutL and MutS functions in activation of MutH. MutH is an endonuclease, it nicks hemimethylated DNA in newly synthesized unmethylated DNA strand and directs mismatch repair. (Jun et al., 2006; Alberts et al., 2002)
In eukaryotes, heterodimers and homodimers of MMR proteins also are involved in mismatch repair and meiotic crossing over (Baker et al., 1996). MLH1 can form heterodimers with PMS1, PMS2 and MLH3. Particularly, the MLH1-MLH3 heterodimer is essential for meiosis, specially for crossover recombination which occurs after double strand breaks (DSB) formation. Also, monomers MSH4 and MSH5 are responsible for the same, whereas PMS1 is responsible for DNA mismatch repair itself and not for crossover recombination . (Baker et al., 1996; Prolla et al., 1994) Defects in the MMR system leads to high mutation rates and cancer risk, both in human and in mice. For example in mice, various neoplasias like lymphoid, intestinal, skin, and other internal organ tumors occur in mutants with MMR defects. In human, particularly germline defects in *Msh2* or *Mlh1*, leads to hereditary nonpolyposis colorectal cancer (HNPCC). (Yao et al., 1999; Edelmann et al., 1999; Loukola et al., 1999). HNPCC is a condition in which susceptibility of developing a colorectal cancer is inherited.

1.7.2 Homologous recombination and its role in DSB repair

1.7.2.1 DSB repair

DSB is the most dangerous type of DNA lesion, as no intact template is present on the broken chromatid to copy and replace the damage. As such, complementary bases on the homologs or sister chromatids can be used as template to copy the DSB region of the damaged chromosome. This type of repair is called homologous recombination (HR) (Baker et al., 1996; Moynahan and Jasin, 2010). A major role of homologous recombination in meiosis as well as mitosis is double strand break repair (Baker et al., 1996; Moynahan and Jasin, 2010). An alternative to HR is Nonhomologous end joining (NHEJ) where broken ends are simply rejoined by DNA ligation. In higher eukaryotes, NHEJ mediated DSB repair is more common in somatic cells (Srivastava and Raman, 2007; Wyman et al., 2004). HR is more complex and time consuming than NHEJ but is more accurate, as no DNA sequence is lost during the repair. According to observations made by Srivastava et. al., HR-mediated repair activity is observed in all major stages of spermatogenesis, showing highest level in spermatocytes (Srivastava and Raman, 2007). Besides HR and NHEJ, the single-strand annealing (SAA) pathway can also be used for DSB repair (Wyman et al., 2004).
In meiosis, HR mediated repair starts with endonuclease cleaving the flanking regions from the point of DSB towards 3' single strand overhang (Figure 1.5). Then, the overhang of unpaired ssDNA searches for complementary base pairs in the homolog and strand invasion and strand extension takes place, thereby synthesizing correct DNA sequence for the gap in the DSB region. In some cases a four DNA stranded structure known as a double Holliday junction is formed, this lead to formation of crossover (i.e. reciprocal exchange of the chromosome arms that flank the break) giving rise to genetic variations. In other cases, chromosomes without crossover (i.e. no exchange of flanking arms) are formed. (Alberts et al., 2002; Wyman et al., 2004; Kauppi et al., 2004).

MMR proteins also have role in meiotic recombination and are involved in various aspects of recombination (Baker et al., 1996). Inactivation of MMR pathway increases mutation rate and susceptibility to tumor development (Modrich and Lahue, 1996). Particularly MLH1 function is crucial in early stage of meiotic cell division, specifically in HR. MLH1 proteins localizes to sites of crossing over during meiotic cell division and is responsible for chiasma formation (Edelmann et al., 1996; Yao et al., 1999; Baker et al., 1996).

To highlight role of MLH1 in HR, MLH1 interacts with multiple proteins and complexes involved in replication. MLH1 protein is found to associate with MRE11-RAD50-NBS1 (MRN) complex, MRE11 and Rad51. MRN complex is associated with various function in HR. MRE11 protein functions as ds-DNA exonuclease and is also involved in DNA-binding and ssDNA annealing activity. Rad51 has DNA strand exchange, ATP-dependent DNA-binding, and ATP-dependent homologous pairing activities. (Wyman et al., 2004)

1.7.2.2 Homologous recombination in meiosis

The exchange of genetic material between homologous DNA sequence is known as homologous recombination (HR) or general recombination. The frequency of HR can range from one to six crossing over per pair of homologous chromosomes, which is organism specific. HR is of great importance as it is responsible for the assortment of chromosomes in gametes, it brings about variation in genome and also has a major role in DSBs repair. (Baker et al., 1996; Hochwagen and Marais, 2010)
HR is responsible for reshuffling of DNA sequence. These rearrangements bring upon genetic variation, which leads to evolution of organism in long term (Hochwagen and Marais, 2010). In meiosis, a crucial step of gametogenesis is segregation of haploid set of chromosomes into each gamete, this involves synapsis and crossing over of homologous chromosomes (Baker et al., 1996; Hochwagen and Marais, 2010). This segregation also allows proper alignment of homologous chromosomes during metaphase of meiosis I (Hochwagen and Marais, 2010). Without these events various abnormalities occur such as aneuploidy and even cell cycle arrest (Baker et al., 1996).
Meiotic recombination events are clustered into narrow (1-2 kb) segments of the genome called hotspots (Jeffreys et al., 2004; Goldstein, 2001). These hotspots are flanked on both sides by large regions of recombinationally suppressed DNA, where recombination is rarely observed (Hochwagen and Marais, 2010; Kauppi et al., 2004).

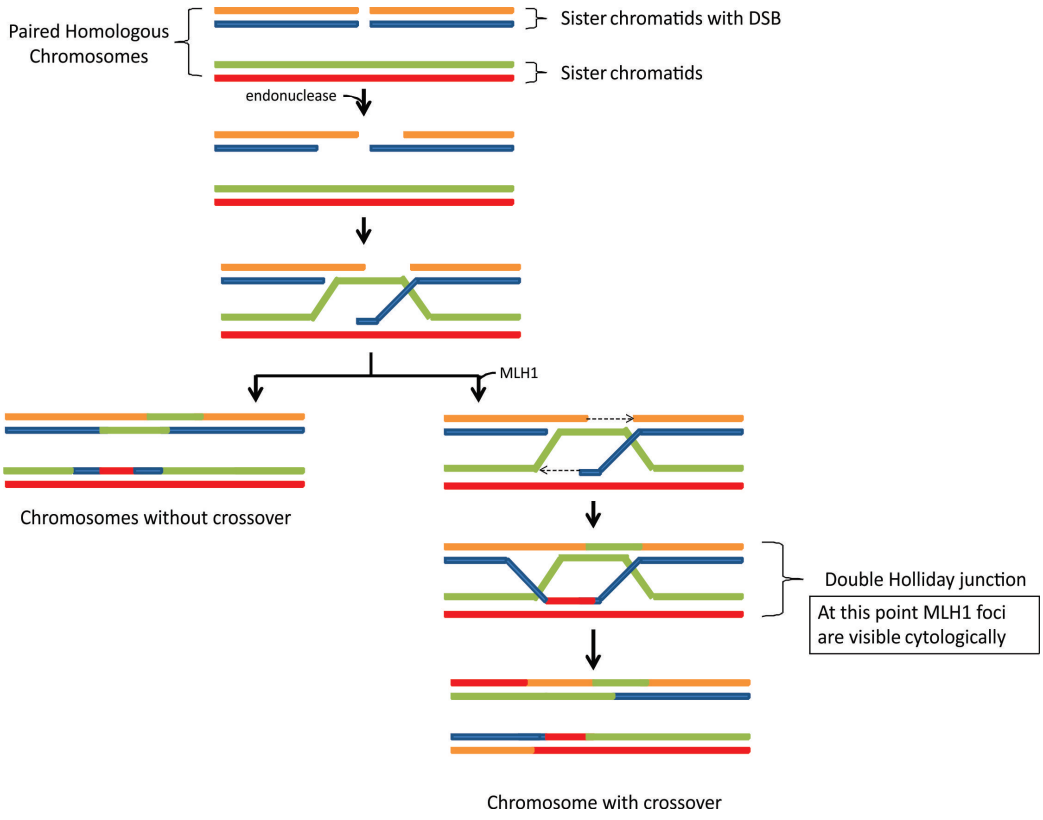

Figure 1.5: Schematic of HR in meiosis

1.8 $Mlh1^{-/-}$ mouse at a glance

The first mice carrying a null mutation in the $Mlh1$ gene was generated by Edelmann et.al. by deleting a portion of the 5' end of the *Mlh1* gene in mouse embryonic stem cells (Edelmann et al., 1999). Both homozygous and heterozygous deletant mutant mice showed susceptibility to tumor development in HNPCC. 32% of the $Mlh1^{+/-}$ mice and 72 % $Mlh1^{-/-}$ mice developed tumor (Edelmann et al., 1999).
In mouse, the Mlh1 gene is present on chromosome 9 (Edelmann et al., 1996).Deletion mutant mice are sterile; both male and female mouse show this phenotype. In male mice, spermatocytes cannot proceed beyond Metaphase stage of Meiosis I due to lack of crossovers, and subsequent chromosomes to align correctly onto the spindle apparatus. This leads to apoptosis of the spermatocytes, presumably due to action of spindle assembly checkpoint proteins (Eaker et al., 2002).
Also, testes of *Mlh1* mice are approximately half the size of wild type and no mature spermatozoa are formed. In female mice, although there is production of oocytes and normal mating, fertilized oocytes cannot develop beyond the single-cell stage. (Edelmann et al., 1996)

1.9 Microsatellite instability(MSI) in the absence of functional MLH1

Loss of MLH1, PMS2 and MSH2 function leads to MSI in somatic and meiotic tissue in both mice and human, relatively in high frequency (Yao et al., 1999; Baker et al., 1996; Edelmann et al., 1999). In yeast, mutation in Mlh1, Msh2, Pms1 results in higher degree of MSI than mutation in Msh4 and Msh6 (Alver et al., 2013). MMR defects are observed in many human malignancies and can be directly associated to MSI (Gutmann et al., 2003; Loukola et al., 1999; Edelmann et al., 1999). This MSI is a consistent feature in HNPCC, as such, germline MSI analysis can be used for prescreening HNPCC (Loukola et al., 1999; Aaltonen et al., 1993). Mlh1 mutation is a major cause for MSI in HNPCC patients (Loukola et al., 1999; Gutmann et al., 2003; Edelmann et al., 1999), according to a study ,79% (65/82) of mutation-positive HNPCC families had *Mlh1* mutation and MSI screening detected 83% of these mutations accurately (Loukola et al., 1999). In Single molecule PCR study performed by Baker et.al. at characterized microsatellite loci in spermatocytes of $Mlh1^{-/-}$ mice , MSI was observed in 14% (32 out of 229) of microsatellite locus *D9Mit67* and 18.3% (33 out of 180) of microsatellite lo-

cus *D1Mit355* and for the somatic samples mutation frequency was found to be 19 % per and 20.3% per somatic cell for respective locus(Baker et al., 1996).

Not much is known about minisatellite instability in absence of functional MLH1 protein. An study conducted by Alver et.al. in yeast showed minisatellite instability to be higher in $Mlh1^{-/-}$ strain than in wild type. They came to this conclusion by quantifying blebbling in stationary phase yeast cells. In $Mlh1^{-/-}$ yeast strain 10 blebs per colony were observed, compared to two in wild type strain (Alver et al., 2013). Further, there has not been any study focusing on germline minisatellite instability in presence and absence of MLH1 protein in mouse.

1.10 Aims of the study

The goal of the project was to:

1. Optimize pipeline for the analysis of germline minisatellite instability.
2. Discover novel minisatellites which contain potential DNA binding motifs for PRDM9 with G-quadruplex potential.
3. Study germline minisatellite instability in the presence and absence of MLH1 protein.
4. Study somatic minisatellite instability in the presence and absence of MLH1 protein.
5. Compare germline and somatic minisatellite instability results from this project with published microsatellite instability (MSI) in $Mlh1^{-/-}$ mouse.
6. Compare germline and somatic minisatellite instability results from this project with published, characterized minisatellites in wild type mouse.
7. DNA binding motif prediction for native and mutant PRDM9 ZF.
8. Compare predicted DNA binding affinity of native and mutant PRDM9 ZF to native PRDM9's DNA binding motif.
9. Compare predicted DNA-protein binding affinity of *Prdm9* alleles to minisatellite *Ms-X165.369*.

Chapter 2

Materials and Methods

This chapter covers the experimental design with the main steps of the project. Starting materials were testis and somatic tissues collected from mutant and wild type C57BL/6J strain mice. Naming of different samples is described in Table 2.1. The work-flow of the project can be seen in Figure 2.1. Each section of this chapter describes individual steps of the analysis along with its aim and importance within the project.

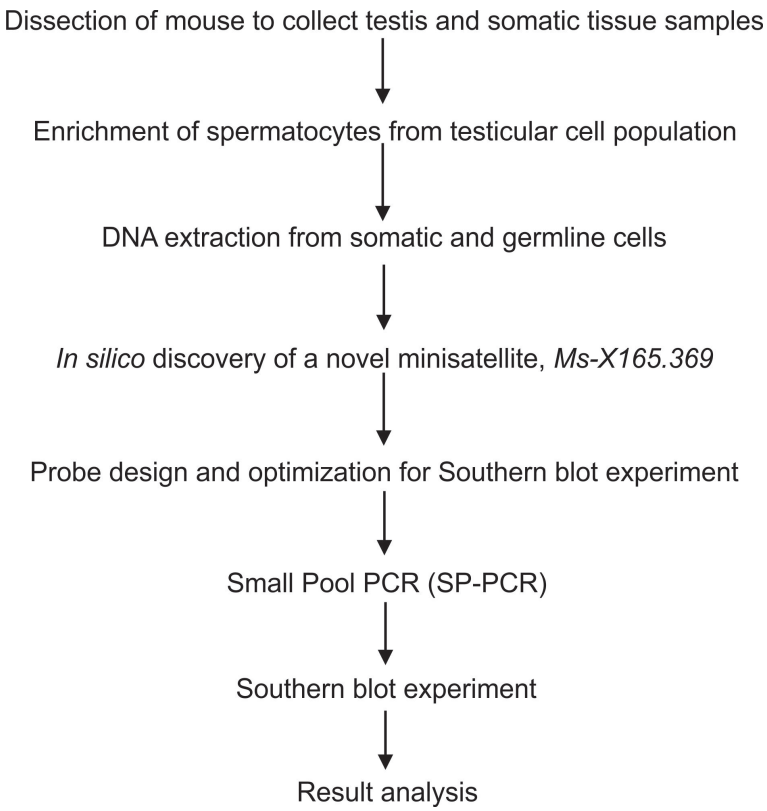

Figure 2.1: Work-flow of the project

Sample name	Origin (DNA extracted from)
$Mlh1^{-/-}$ spermatocyte	enriched spermatocyte population from testis of *Mlh1* knockout mouse strain
$Mlh1^{+/+}$ spermatocyte	enriched spermatocyte population from testis of $Mlh1^{+/+}$ mouse strain
$Mlh1^{-/-}$ somatic	somatic tissue (kidney) of *Mlh1* knockout mouse strain
WT somatic	somatic tissue (Kidney) of wild type inbred mouse

Table 2.1: Sample name and origin

2.1 Dissection of mice to recover testis and somatic tissue samples

This step was conducted to get the starting materials for the project.

Eight weeks old *Mlh1* knockout, $Mlh1^{+/+}$ and wild type mice were sacrificed using CO_2 (Carbon dioxide), then cervical dislocation. Testis and kidneys were dissected by making an incision in the abdominal area. Testes are present on either side of the urinary bladder attached to the ends of adipose tissue(fat pads). Fat pads were pulled out using sterile forceps and testes were removed by chopping it off from the fat pads. The samples were collected in Eppendorf tubes, transferred to the laboratory on dry ice and stored in -80 °C until further use.

2.2 Enrichment of spermatocytes from the testicular cell population

This step was performed to collect an enriched cell population of spermatocytes for further analysis. A bigger cell pellet obtained in the end of the step, indicated a higher yield of cells.

Tunica albuginea were remove from each testes. The testes were transferred to a 15 ml tube containing 2 ml of Testes cell isolation medium (TIM) (104 mM NaCl, 45mM KCl, 1.2 mM $MgSO_4$, 0.6 mM KH_2PO_4, 0.1% (w/v) glucose, 6 mM Na Lactate, 1 mM Na pyruvate, pH to 7.3, filter sterilized at RT). 200μl of freshly prepared collagenase solution was added (20 mg/ml collagenase (Gibco) in TIM) to the tube and was shaken in a Thermomixer (Eppendorf) for 55 minutes at 32°C and 600 RPM with occasionally inverting

of the tube. As the first washing step, TIM was added to a total volume of 15 ml, seminiferous tubules were allowed to settle for three minutes and the supernatant was removed. The washing step was performed twice. Tubules were resuspended in 2 ml TIM. 200 μl of Freshly prepared Trypsin solution (7 mg/ml of Trypsin (Sigma T9935) in TIM) was added to the tube, followed by 20 μl of DNase I solution (400 μg/ml DNase I (Roche) in TIM). The tube was shaken at 600 RPM for 15 minutes at 32°C with occasionally inverting of the tube. 500 μl of freshly prepared Trypsin Inhibitor solution (20 mg/ml of Trypsin Inhibitor (Sigma T9003) in TIM) and 50 μl of 400 μg/ml DNase I solution was added. Repeated up and down pipetting was done for two minutes to separate the cells completely. The cell suspension was filtered through monofilm nylon filter of 10 μm pore size (Partec CellTrics), 200 μl at a time. The filter membrane was rinsed with 500 μl of PBS into 1.5 ml Eppendorf tube. For recovery of remaining cells, the filter membrane was placed in the Eppendorf tube used in the previous step, and PBS was added to make total volume 1 ml. The tube was incubated for 30 minutes at 32°C with gentle shaking. The filter membrane was removed and discarded. Heat inactivation of DNase I was performed by incubating this Eppendorf tube for 10 minutes at 75°C. The cell suspension was centrifuged at maximum speed for 15 minutes, and the very small pellets were used for genomic DNA extraction.
To check the size and density of the cells, 10 μl of cell suspension per sample was transferred onto a glass slide and 5 μl of DAPI was added. Cells were visualized under an epifluorescence microscope.

2.3 Obtaining cells from somatic tissue samples

This step was performed to recover as many cells as possible from the somatic tissues. Similar to previous step, a bigger cell pellet obtained in the end of the step, indicated a higher yield of cells.

Each tissue sample was defrosted at room temperature. Using a razor blade, the defrosted sample was chopped into small pieces in a weigh boat, then rinsed with 1-2 ml of PBS into a 15 ml falcon tube. The tube was centrifuged at 4000 rpm for 1 minute. The supernatant was discarded, leaving the pellet behind. The pellet was resuspended in 1 ml 1X SSC and centrifuged at 4000 rpm for 1 minute. The supernatant was discarded, and the pellet was again resuspended in 1X SSC making the total volume 1 ml. The

cell suspension was transferred into a 1.5 ml Eppendorf tube. The tube was centrifuged for approximately 90 seconds at 903 g (3500 rpm). Time for centrifuge was adjusted such that the pellet was neither too packed, nor too loose. The supernatant was discarded and from the cell pellets, DNA was extracted as described in next step.

2.4 DNA extraction from somatic and germline cells

This step was performed to obtain DNA of high quality for subsequent analysis.

To each tube with a cell pellet obtained from the previous two steps, 300 μl 1 X SSC was added. The pellet was resuspend by vortexing. After that, 35 μl of 10% SDS was added, followed by 10 μl of Proteinase K (20 mg/ml) and then 40 μl of 2-mercaptoethanol was added. Tube was thoroughly mixed by hand and was incubated at 37°C for 1 hour.
To extract DNA, phenol chloroform DNA extraction was performed. 200μl phenol and 200μl chloroform were added to each Eppendorf tube and mixed vigorously by hand. The tubes were centrifuged for 2 minutes at full speed. The upper (aqueous) layer was transferred into a new 2 ml tube, and the lower (organic) layer was discarded. Caution was taken not to pipette the interphase layer. To each tube, 1/10 volume of 3M NaAc was added, followed by 2× volume of ice-cold absolute ethanol and mixed thoroughly by inversion. DNA was allowed to precipitate for 10 minutes at -80°C. The tube was centrifuged for 1 minute at maximum speed. The supernatant was discarded retaining the precipitated DNA. 1ml cold 80% ethanol was added to wash the pellet and was centrifuged for 2 minutes at full speed. The supernatant was discarded very carefully leaving 150 to 200 μl behind and centrifuged again for 2 to 3 minutes. Rest of the supernatant was pipetted out. The DNA pellet was air-dried at room temperature for a few minutes until ethanol had completely evaporated. The pellet were dissolved in 50-60μl of 5mM Tris-HCl by letting it stand for 2 to 3 minutes, then by gentle pipetting. DNA concentration was measured using Qubit dsDNA HS Assay Kit (Life Technologies) and Qubit dsDNA BR Assay kit (Life Technologies). Quality and concentration of DNA were also checked by running an aliquot of each DNA sample on a 1% agarose gel.

2.5 *In silico* discovery of a novel minisatellite, *Ms-X165.369*

The aim of this step was to find previously undescribed minisatellites with G-quadruplex forming potential, that could also function as potential DNA binding sites for PRDM9 ZFs. G-quadruplex are G-rich nucleic acid sequences that can readily form four-stranded structures. G-quadruplexes have a role in chromatin packaging, recombination and CpG methylation (Kumar et al., 2011; Bochman et al., 2012).Additionally, minisatellites with this potential are highly unstable. Also, feasibility for further analysis in wet-lab was considered while making the selection.

For this part of the study, we started analysis from far subteromeric end of the X chromosome towards the centromere. We decided to begin form subtelomeric end because there were high chances of finding PRDM9 binding sites at this region. The subtelomeric region shares boundary with Pseudoautosomal region(PAR). PAR is a small region of homology between X and Y chromosomes which has high-frequency of crossing over. These hotspots of crossing over should be specified by PRDM9, as such the subtelomeric region should harbor potential DNA binding motifs for PRDM9.
These regions were analyzed for the presence of minisatellites which had G-quadruplex forming potential and contained a potential PRDM9 binding motif. First, to obtain the minisatellites in this region, tandem repeats were identified using the web based tool Tandem Repeat Finder (tandem.bu.edu/trf/ trf.html) (Benson, 1999). The output was filtered with three criteria for minisatellite: repeat unit size of 10 to 104, Number of repeats seven or above and GC content greater than 50 %. Another round of filtering was performed using QGRS Mapper (bioinformatics.ramapo.edu/QGRS/ index.php) (Kikin et al., 2006) to obtain minisatellites with G-quadruplex forming potential. Out of the minisatellites having G-quadruplex potential, five were selected whose instability could be feasibly analyzed in the wet lab. Theoretically, in these five selected minisatellites, a 10% size shift in total product should be detectable by agarose gel electrophoresis. For example, for an minisatellite of 100 bp length and each repeat unit of 10 bp length, a gain or loss of a single repeat unit will be increase or decrease of 10 bp, which is 10 % of the total minisatellite length. These five were tested for its potential as a PRDM9 motif using the web-based tool "A DNA binding site predictor for Cys2His2 Zinc Finger Proteins (http://zf.princeton.edu/index.php)" (Persikov et al., 2009; Persikov and Singh, 2014). This tool predicts zinc finger binding using either linear or polynomial Support Vector Machines

(SVMs).The authors have stated polynomial SVM performs better for this prediction, and we have used the same for our analysis (Persikov and Singh, 2014).
This program uses SVM as a classifier for DNA-protein binding prediction. A positive score signifies DNA-protein binding and negative score signifies no binding of protein to DNA, in our case binding of PRDM9 zinc fingers to the five minisatellites with G-quadruplex potential. The magnitude of the score is proportional to the relative DNA-protein binding affinity, higher SVM score predicts more affinity.
For the wet lab, primers were designed for the minisatellites fulfilling all the filtering criteria mentioned above, and then optimized. The PCR products were visualized by running on an agarose gel for 45 minutes at 120 V. The most optimized of all was selected for further analysis.

2.6 Probe design and optimization for Southern blot experiment

This step was performed to design probe and optimize conditions for probe preparation. A probe in Southern blot experiment is a specific segment of ssDNA that is complementary to a desired DNA sequence, crosslinked to a nylon membrane. In this project we have used radio-labeled, denatured dsDNA amplicon (PCR product)containing respective minisatellites, along with flanking regions, as probe. Any PCR requires a primer pair and an optimized PCR program. An optimal primer must have a length ranging from 18 to 25bp and should not form self or cross dimer. Also, the melting temperature (Tm) difference between forward and reverse primers must not be more than three degrees and GC content must be more than 50%.
A probe must be highly specific such that it hybridizes only with the desired DNA sequence crosslinked to the membrane. Thus, probe design and optimization is at the heart of any Southern blot experiment. The most optimal length of a probe is from 100 to 1000 bp, but the range can be can be from 25 to 2500 bp.

P32-labeled DNA probes were used for the Southern blot experiment. First, for each minisatellite, its sequence along with few kb flanking region was retrieved from the Ensembl database(www.ensembl.org). The sequence was annotated and the expected size of the amplicon was calculated. For *Prdm9* (Buard et al., 2014), *Mms80* (Bois et al., 2002) and *Ms6hm* (Polyzos et al., 2006) previously published primers were used. For *Ms-X165.369*, primers

were designed in the region flanking the minisatellite using the Primer-BLAST (ncbi. nlm.nih.gov/tools/primer-blast) tool. Self and cross dimer, and melting temperature(Tm) were checked using a web-based tool(thermoscientificbio.com/webtools/multipleprimer). The list of primers used and designed in this project is listed in Table A.1. The size of the minisatellite and the individual amplicons is listed in table 2.6. The PCR program for all the primer pairs were optimized using gradient PCR, as well as altering number of cycles, adding separate PCR cycles, and altering extension times where ever required. The 11.1X buffer system used for the PCR reactions was adopted from elsewhere (Jeffreys et al., 1990). Components of the 11.1X buffer can be seen in Table A.2. The PCR master mix component can be viewed in Table A.3. The PCR products were run in 1% agarose gel for 60 minutes to visualized the amplified fragments. Optimal conditions should yield only a single band of expected size. These bands were excised out from the gel. Then DNA from the excised pieces of gel was purified using NucleoSpin Gel and PCR Clean-up kit. Concentration of extracted DNA was measured using the Qubit dsDNA BR Assay kit (Life Technologies). DNA were stored at 4 °C. Details of radio-labeling of the DNA is explained in the next section.

locus name	**Minisatellite length (in bp)**	**amplicon length (in bp)**
Prdm9	924	1636
Ms-X165.369	225	879
Ms6hm	119	382
Mms80	5700	6073

Table 2.2: Size of the minisatellites and the individual amplicon

2.7 Small Pool PCR (SP-PCR)

This step was performed to analyze minisatellite instability in our samples. This step was such designed that each PCR reaction had a small number of amplifiable molecules. After PCR, the PCR products were run for a long period of time in gel to obtain desired resolution, such that even a single loss or gain of repeat could be detected.

Using values obtained from the Qubit measurement as reference concentration , each sample was diluted to five amplifiable DNA molecules per μl. For sample of somatic origin dilutions was made considering each genome to be approximately 6pg. For the sample of germline origin, dilutions was based

on the fact that these samples (i.e. enriched spermatocytes) are in zygotene stage of Prophase I of meiosis I. At this stage a single amplifiable molecule (i.e. single genome) is 4n homologous chromosome. As each 2n chromosome is 6 pg, this 4n structure will be 12 pg. A 50% PCR efficiency factor is added as not all but every second PCR reaction yields a positive PCR amplified product. Thus the final diluted samples had 5 amplifiable molecules(i.e. genomic DNA) per μl which was calculated to correspond to approximately 120 pg and 60 pg for germline and somatic samples respectively.
As the concentration of $Mlh1^{-/-}$ spermatocyte sample was low even when measured using the Qubit dsDNA HS kit (range 0.2-100 ng), the number of molecules was estimated using Poisson approximation by visualizing gel image with 12 SP-PCR reactions. Poisson approximation can be mathematically represented as:

Number of amplifiable molecule = - ln (number of negative PCR reactions/ total number of PCR reactions)

As working with extremely low concentration of DNA, extra precaution was taken to reduce contamination and DNA loss in handling and storing the samples. Low concentration DNA samples were stored in 5 ng/μl herring sperm solution and also all the dilutions were performed in presence of 5 ng/μl herring sperm. Herring sperm is a carrier DNA and reduces loss of sample DNA due to its sticky nature. Also low retention pipette tips (Sartorius) and tubes (DNA LoBind Tubes) were used because of the same reason. In order to minimize risk of PCR product contamination, all the solutions and PCR preparations were performed in a UV cabinet UVT-B-AR PCR cabinet (Grant), UV treated for 15 minutes before use.
The same primers (Table A.1), buffer (Table A.2), master mix (Table A.3), and respective optimized PCR programs (Table 3.3 - 3.5) as for the probe design (section 2.6) were used for SP-PCR for each locus. As template DNA, 1 μl of the sample (approximately equal to 5 DNA molecules) was used. Additionally, 1.2 μl of 5ng/ μl herring sperm (Invitrogen) was added to each reaction making total volume of PCR reaction to 10 μl.
The PCR products were run on a 40 cm long, 0.7% agarose gel (2.45 mg of agarose in 350 ml of 1X TBE with 1 μl/100 ml Etbr) at 120 V for 17 to 24 hours in 3500 ml of 1X TBE buffer. Horizontal electrophoresis tank (C.B.S scientific) was used for this purpose. To each PCR product, 1X loading dye with 5μl/ml EtBr was added before loading samples into the wells. 1.5 μl of 1kb DNA ladder (GeneRuler ThermoScientific) and Lambda/ Hind III (Invitrogen) marker were also run along the samples. The DNA fragments were visualized under UV and portion of gel with these fragments was cut.

Southern blot experiment were carried on for these pieces of gels.

2.8 Southern blot experiment

This step was performed to visualize the SP-PCR results from previous step. The SP-PCR products were very hard to visualize in a gel as our concentrations were lower than detectable limits of a gel. Thus a method with higher detection sensitivity, such as Southern blot, was required to visualize the SP-PCR results.

The cut portions of the gel from previous step were removed from the casting tray and cross-linked by UV exposure at 180 J/cm^2. A small piece of gel from the upper left corner was cut to mark orientation of the gel. To start the transfer, the gel was washed twice with denaturation buffer (1.5 M NaCl/0.5 M NaOH) for 15 minutes each with gentle agitation. This step was followed by wash with Neutralizing buffer (1.5M NaCl/1M Tris-HCl,pH 7.0) twice, first for 10 minutes and then for 20 minutes with gentle agitation. Transfer of the DNA from the gel to a nylon membrane (GE Healthcare Amersham Hybond-N+) was performed overnight in setup as shown in Figure 2.2.
On the next day, the nylon membrane was dried in the oven at 80°C for 15 minutes, followed by 180 J/cm^2 UV exposure. The membrane was stored in a dry and clean envelope.
For Southern blot hybridization, the membrane was transferred into a hybridization bottle with DNA facing inward. For *Prdm9*, all the membranes were hybridized in the same bottle by sandwiching them in between sheets of hybridization mesh. Pre-hybridization was performed for one hour at 65°C in a hybridization oven (Techne) with approximately 40 ml prehybridization buffer (7% SDS, 0.5 M phosphate buffer, 1mM EDTA).
To label the DNA to be used as probe, 25 ng of template DNA along with 5 ng of ladder was combined in an Eppendorf tube, final volume of 12 μl was made by adding water. The DNA was denatured at 100°C for 10 minutes and then quickly transferred into ice. After denaturation, 3 μl of dNTP (without dCTP) mix (Roche), 2 μl reaction mixture (Random Primed DNA Labeling Kit,Roche), 2 μl of dCTP[$\alpha -^{32} P$] (800 $Ci/mmol$) (EasyTides) and 1 μl Klenow enzyme (Random Primed DNA Labeling Kit,Roche) was added to the tube and incubated at 37 °C for 30 minutes. 2μl of 0.2 M EDTA (pH 8.0) was added to stop the labeling reaction, also 3 μl of 1 mg/ ml Herring Sperm (Invitrogen) was added, which functions as carrier DNA. To remove unincorporated radionucleotides, the reaction mix was then transferred to Quick spin columns (Roche), and centrifuged for 4 minutes at 1000 G. A de-

naturation step for probe was performed way heating the tube with labeled probe for 5 mins at 100°C. Thus, the probe was ready to use.
Pre-hybridization buffer was poured out of hybridization bottle, and 40 ml of fresh hybridization buffer (same as pre-hybridization buffer) was added to the tube, then the probe was added. Hybridization was performed by incubating the bottle overnight at 65 °C in hybridization oven with gentle rotation. Next day, the membrane was washed five times with wash buffer (1% SDS, 40 mM phosphate buffer, 1 mM EDTA). The first wash was just rinsing the membrane, next four washes were of 15 minute each at 65 °C. The membrane was transferred to a GE Healthcare storage Phosphor image screen and was exposed overnight, next day radioactivity in the screen was detected using molecular imager (GE Amersham Molecular Dynamics Typhoon 9410). The images obtained were used for further analysis.

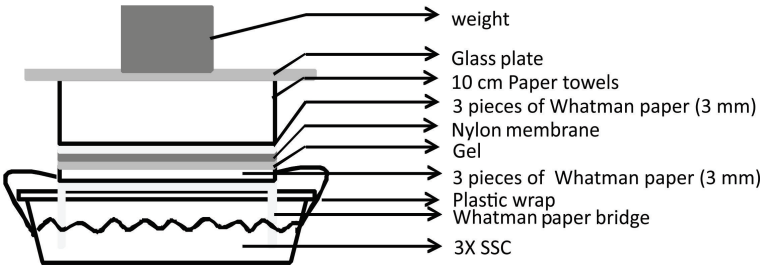

Figure 2.2: Southern transfer setup

Chapter 3

Results

In this chapter we will show the important results obtained from various methodological steps discussed in previous chapter.

3.1 Testes recovered from the *Mlh1* knockout and $Mlh1^{+/+}$ mouse strain

The size of testis of *Mlh1* knockout was almost half of the $Mlh1^{+/+}$ mouse strain (Figure 3.1).

Figure 3.1: Dissected testes for $Mlh1^{-/-}$ and $Mlh1^{+/+}$ mouse

3.2 Enriched spermatocytes cells from testicular cell population

Filtering the cell suspension through monofilm nylon filters of 10 μm pore size yielded an enriched spermatocyte cell population, majority being desired primary spermatocytes, on the filter membrane. These cells retained in the filter membrane are greater than 10 μm in diameter and cells in the flow-through are less than 10 μm. Almost all the cells collected from the filter were big and round in morphology, characteristic of a spermatocytes. As shown in microscopy images in Figure 3.2, number of cells collected from both the filter and flow-through were very few in number in $Mlh1^{-/-}$ spermatocyte sample compared to $Mlh1^{+/+}$ spermatocyte samples.
For $Mlh1^{-/-}$ spermatocyte samples, cells collected from both filter and flow-though were scarcely populated. More cells were seen in the cell suspension collected as flow-through. For cells collected from the filter, in a given field of view at 10 X magnification (Figure 3.2a) only two cells were visible. In most field of view no cells were observed . No spermatocytes or mature sperm cells were observed in the flow-through.
For $Mlh1^{+/+}$ spermatocyte samples, higher number of cells were seen in both cell suspensions, i.e. collected from filter and from flow through. Almost all of the cells in the flow-through were sperm in different stages, characterized by its hook shaped sperm head. Spermatids and mature sperm cells were completely removed as flow-through.

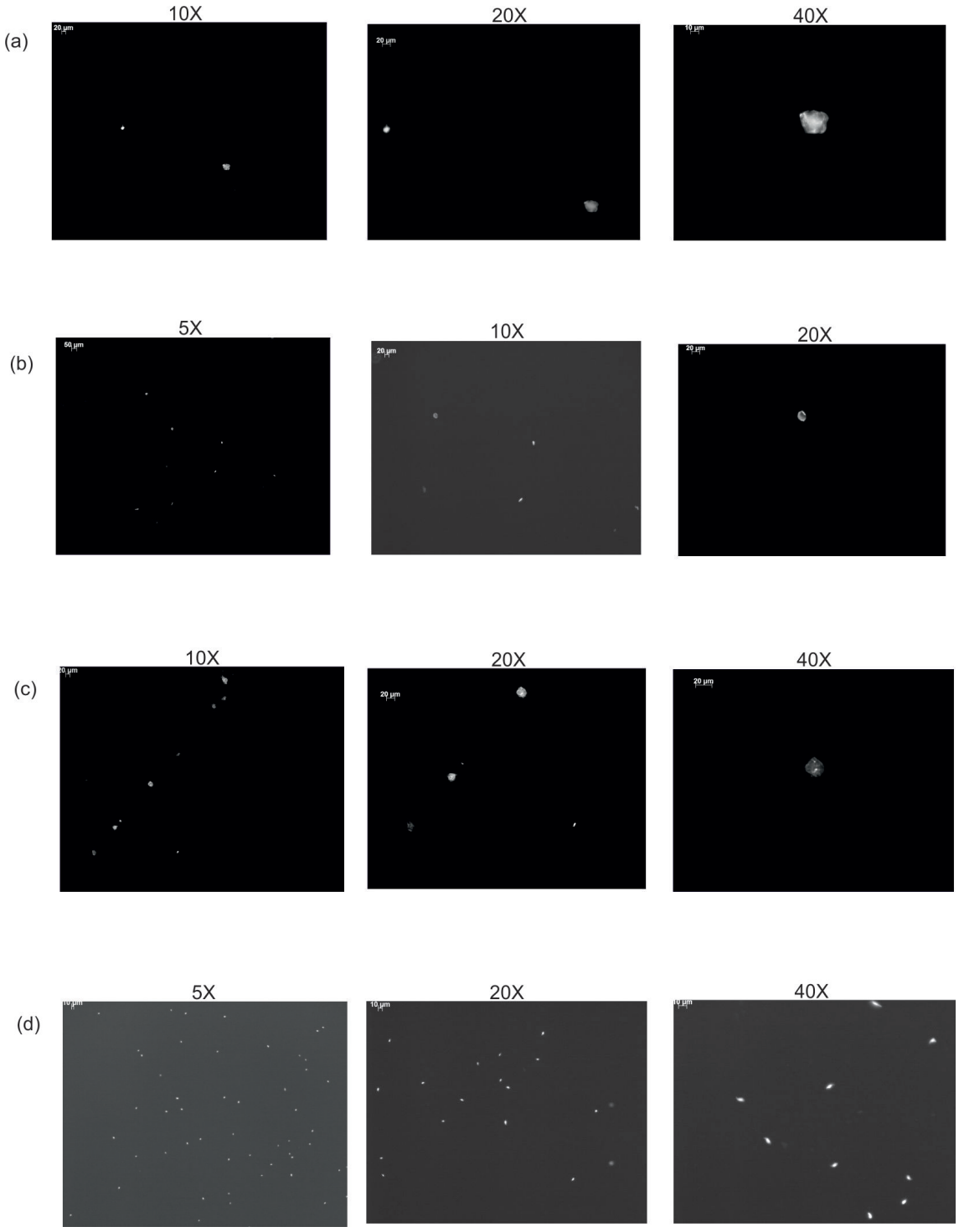

Figure 3.2: DAPI stained microscopic images taken after enrichment of spermatocytes from testicular cell population: (a)view of cells collected from filter for *Mlh* -/- samples, (b)view of cells in the flowthrough for $Mlh1^{-/-}$ spermatocyte samples, (c)view of cells collected from filter for $Mlh1^{+/+}$ spermatocyte samples, (d)view of cells in the flowthrough for $Mlh1^{-/-}$ spermatocyte samples

3.3 *In silico* discovery of novel minisatellite *Ms-X165.369*

In the region analyzed, 40 minisatellite were found using Tandem Repeat Finder. Another round of filtering using QGRS Mapper filtered 12 of these 40 minisatellite as possible G-quadruplex forming minisatellites. Out of 12, five were selected for reasons mentioned in section 2.5. The five selected minisatellites are listed in table 3.1 along with their coordinates on X chromosome, core repeat length, number of repeats, CG% and total length. The coordinates on X chromosome is based on Genome assembly: $GRCm38(GCA\_000001635.4)$. PRDM9 zinc finger domain from progenitor allele was predicted to bind to all five minisatellites. Table 3.2 lists the five minisatellites along SVM scores. Among the five minisatellites, *Ms-X165.369* had the highest SVM score and also showed the best optimized result in the wet-lab experiments. Hence it was selected for further analysis.
Minisatellite *Ms-X165.369* is located at the far subteromeric end of the X chromosome, approximately 4 Mb from the PAR boundary. Repeat unit of minisatellite consists of a 25mer sequence AGAG GGAGAGGGAGGAGGGA-GAAAG with few SNPs, and total array consist of nine such tandem repeats.

Minisatellite locus	**Coordinates on X chromosome**	**core repeat length**	**no. of repeats**	**GC%**	**Total length**
Ms-X165.369	165369082-165369317	25	8.9	59.6	235
Ms-X166.499	166499080-166499505	34	13.7	63.6	425
Ms-X169.597	169597462-169598192	104	7	58.2	730
Ms-X068.848	68848532-68848709	25	7.2	69.7	177
Ms-X152.591	152591310- 152592210	74	12.2	56.7	900

Table 3.1: Five selected minisatellites with G-quadruplex potential.

Minisatellite locus	**SVM score**
Ms-X165.369	7.87
Ms-X152.591	7.84
Ms-X166.499	4.94
Ms-X169.597	4.74
Ms-X068.848	0.92

Table 3.2: Predicted binding potential of PRDM9 ZF from progenitor allele to each minisatellite

3.4 Probe design and optimization for Southern blot experiment

Optimization of primers gave a single DNA fragment of the expected size for the loci as shown in Figure 3.3, except for *Ms6hm*. Even, restriction digestion of genomic DNA by enzyme AluI gave three DNA fragment for *Ms6hm* locus Figure A.1. This can either be because of unspecificity of the primers or sequence similarity between the three fragments. The optimized PCR programs for *Prdm9*, *Ms-X165.369* and *Mms80* can be seen in Tables 3.3, 3.4 and 3.5 respectively. Yet to be fully optimized PCR program can be seen in Table 3.5 in Appendix section. As, only single DNA fragment was observed for the three, restriction digestion step was skipped for probe making. The list of restriction enzymes and the cut sites and restricted fragment sizes for each locus are listed in Table A.8 and Figure A.5 in Appendix. Locus *Mms80* could not be further assayed within the timeframe of this project. Due to lack of time and complication faced with primer optimization for *Ms6hm* locus, this locus was also left out and would be one to be analyzed in the future.

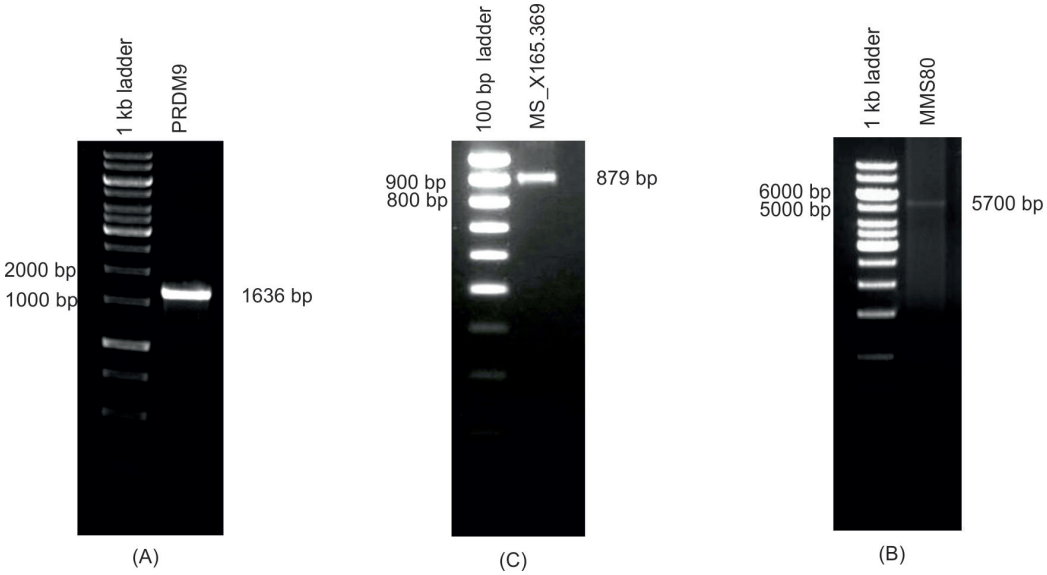

Figure 3.3: Gel image of DNA fragments for (A) *Prdm9*, (B)*Ms-X165.369* and (C)*Mms80*

Temperature	Duration	Cycles
96 °C	30 sec.	X 1
96 °C 58 °C 70 °C	10 sec. 20 sec. 2 min.	X 30
70 °C	7 min.	X 1
4 °C	∞	

Table 3.3: Optimized PCR program for Southern probe making for *Prdm9*

Temperature	Duration	Cycles
95 °C	1 min.	X 1
95 °C 64 °C 68 °C	15 sec. 20 sec. 1 min.	X 3
95 °C 62 °C 68 °C	15 sec. 20 sec. 1 min.	X 32
68 °C	7 min.	X 1
4 °C	∞	

Table 3.4: Optimized PCR program for Southern probe making for *Ms-X165.369*

Temperature	Duration	Cycles
96 °C	100 sec.	X 1
96 °C 61 °C 70 °C	50 sec. 45 sec. 6 min.	X 30
70 °C	7 min.	X 1
4 °C	∞	

Table 3.5: Optimized PCR program for Southern probe making for *Mms80*

3.5 Germline minisatellite instability analyses

Prdm9 and *Ms-X165.369* loci were analyzed for germline minisatellite instability. For *Prdm9*, a total of 151 SP-PCR was performed and in total and 81 of them were PCR positive. And, for *Ms-X165.369*, a total of 36 SP-PCR was performed and 15 of them were PCR positive. Number of SP-PCR performed and number of PCR positive reactions for individual samples can be viewed in Table 3.6 and Table 3.7 for *Prdm9* and *Ms-X165.369* respectively The calculations prior to SP-PCR were done such that each SP-PCR will have five amplifiable molecules, but it was not so for all the SP-PCR reactions. Many SP-PCR reactions did not amplify at all, as shown by number of positive PCRs above. This suggest that some SP-PCRs had even less than one amplifiable molecule. Therefore, number of amplifiable molecules were estimated by Poisson approximation. Poisson approximation of number of amplifiable molecules for each sample in individual experiment(blot)for *Prdm9* is listed in Appendix Table A.7(a) and the overall Poisson approximation for each sample in all the experiments can be seen in Appendix Table A.7(b).For *Ms-X165.369*,Poisson approximation of number of amplifiable molecules for individual samples can be seen in Appendix Table A.7(c).

sample name	Total SP-PCR	Total PCR positive reaction
Mlh1 -/- spermatocyte	29	4
Mlh1 +/+ spermatocyte	31	16
Mlh1 -/- somatic	47	44
WT somatic	44	17
Total	151	81

Table 3.6: Total SP-PCR performed for individual samples along with total positive PCR obtained for *Prdm9*

sample name	**Total SP-PCR**	**Total PCR positive reaction**
Mlh1 -/- spermatocyte	12	1
Mlh1 +/+ spermatocyte	12	2
Mlh1 -/- somatic	12	12
Total	36	15

Table 3.7: Total SP-PCR performed for individual samples along with total positive PCR obtained for *Ms-X165.369*

To visualize the SP-PCR results more accurately, Southern blot experiment was performed right after SP-PCR. The size shifts of the amplicon were observed by using radio-labeled probe, as discussed in the previous chapter.The complete set of Southern blot images are shown below in Figure 3.4,3.5,3.6. An enlarged image of portion in the blots where size shifts were observed are shown in Figure 3.7.

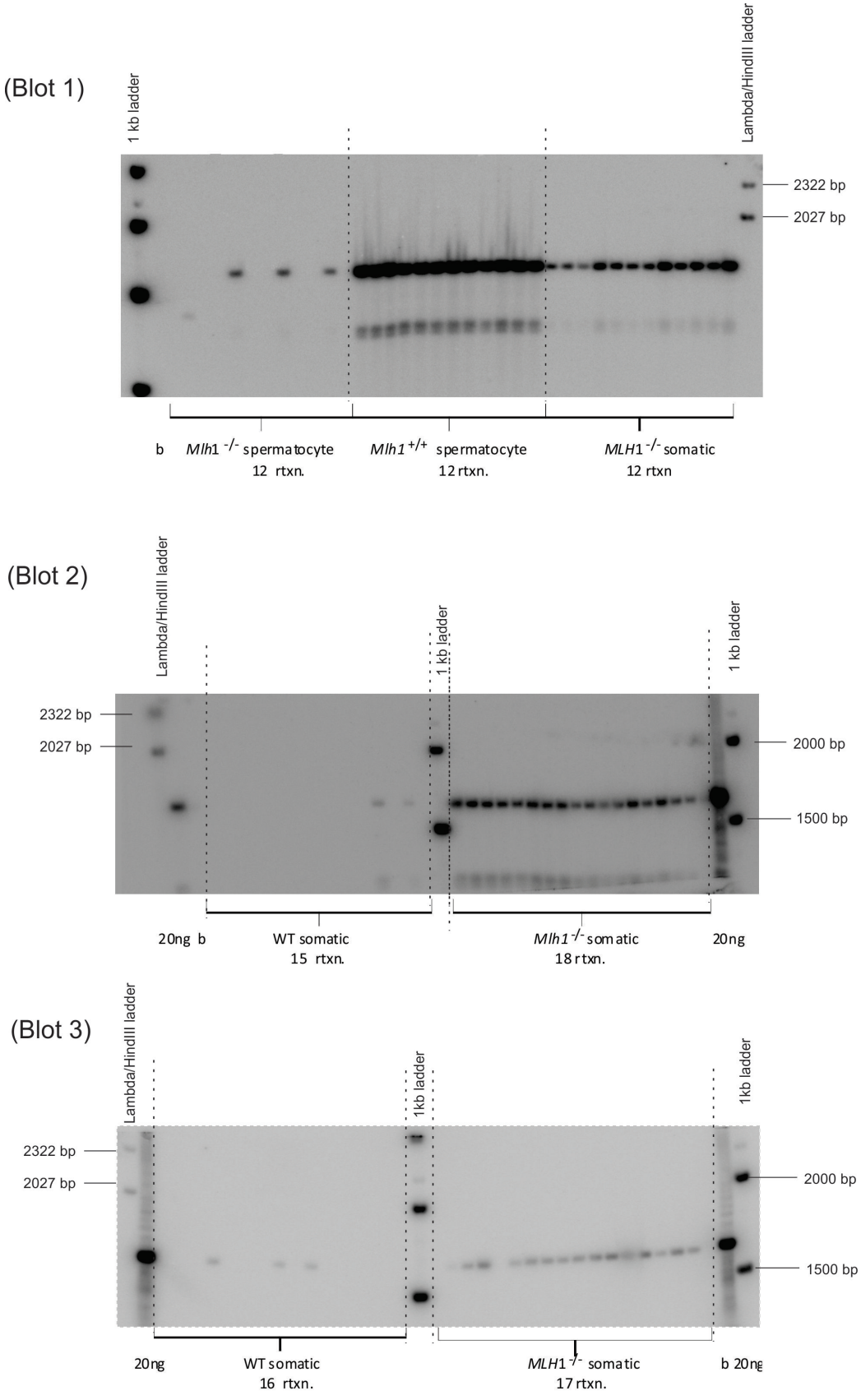

Figure 3.4: Southern blots 1,2 and 3 for *Prdm9* locus

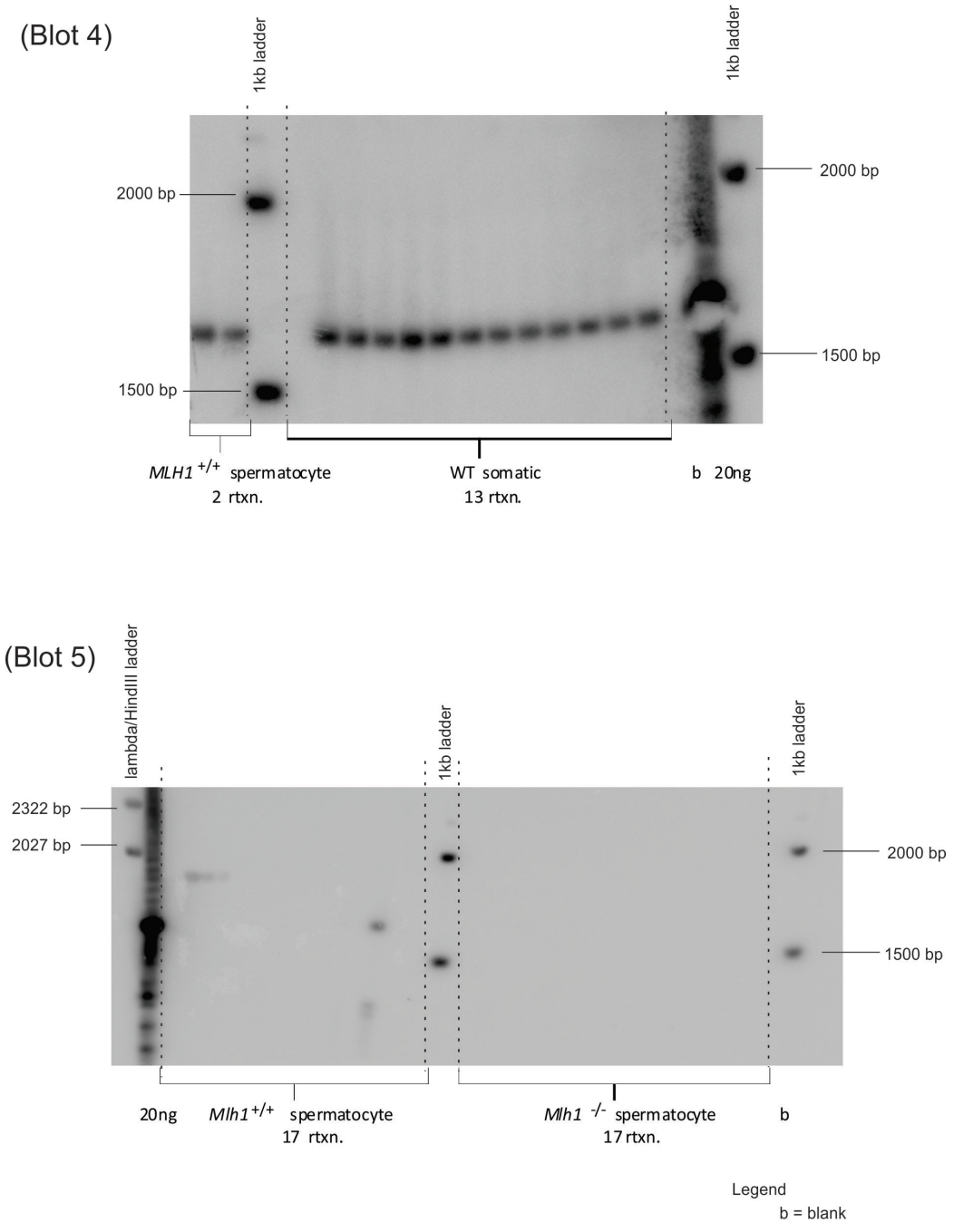

Figure 3.5: Southern blots 4 and 5 for *Prdm9* locus

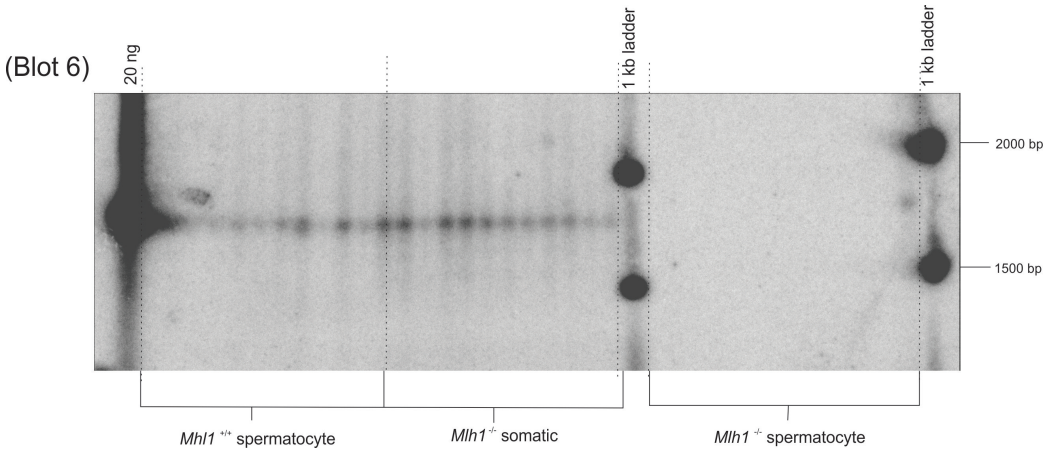

Figure 3.6: Southern blots for *Ms-X165.369* locus

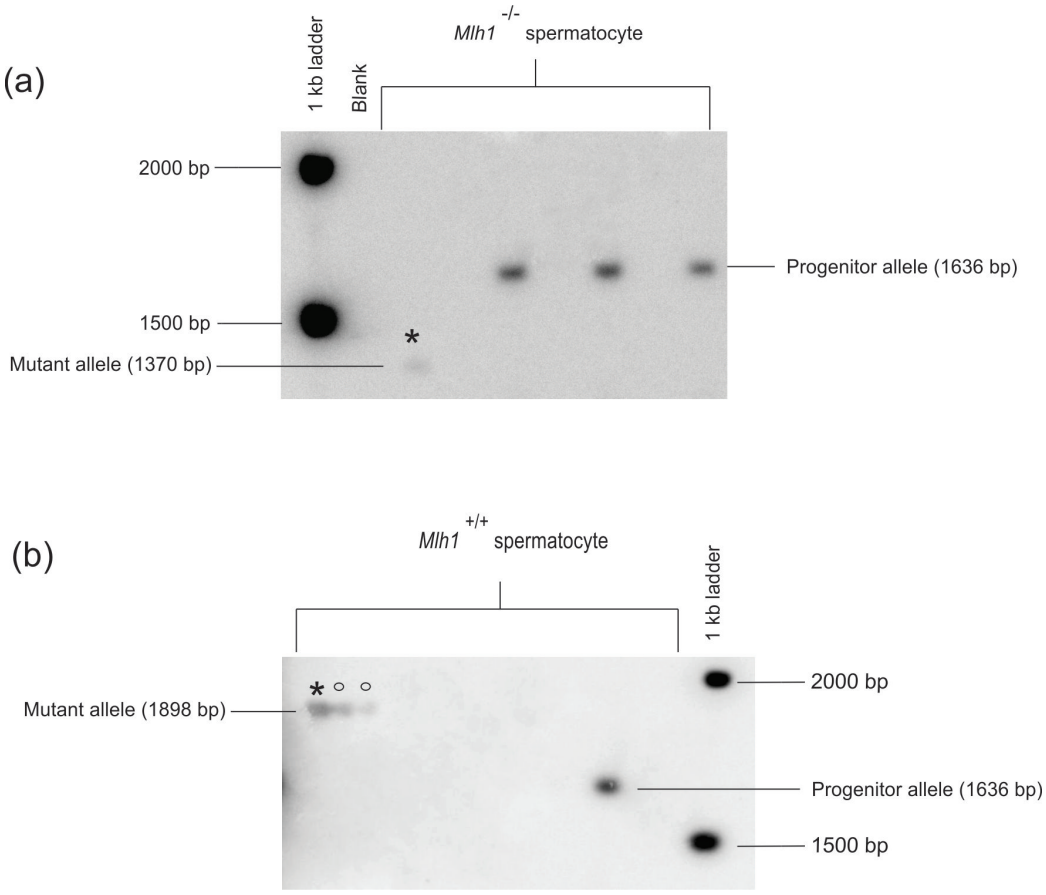

Figure 3.7: Mutant allele of *Prdm9*(indicated by an asterisk)observed in (a)$Mlh1^{-/-}$ spermatocyte sample in blot 1 and (b) $Mlh1^{+/+}$ spermatocyte sample in blot 5. degree symbol in $Mlh1^{+/+}$ spermatocyte sample represents potential artifact.

For accurate measurement of the size shifts myImageAnalysis v1.1 (ThermoScientific) program was used. The skewness of migration of the band, if present, was also corrected with this program. Screen shots of the standard curve for measurement of migration of some interesting mutant alleles can be seen in Appendix Figure A.2, A.3.

The sizes of the progenitor alleles at the *Prdm9* and *Ms-X165.369* locus along with the flanking regions was designed to be 1636 bp and 879 bp, respectively. All the progenitor bands for *Prdm9* were nearly 1636 bp with slight deviations as shown in Figure 3.8 and Appendix Table A.5. Also, for *Ms-X165.369* all the progenitor bands were nearly 879 bp with slight deviation as shown in Figure 3.9 and Appendix table A.6. These deviations can be due to technical and analytical errors like nonuniform current flow through the buffer in electrophoresis tank, and uneven polymerization of gel. But the deviation did not affect minisatellite instability analysis.

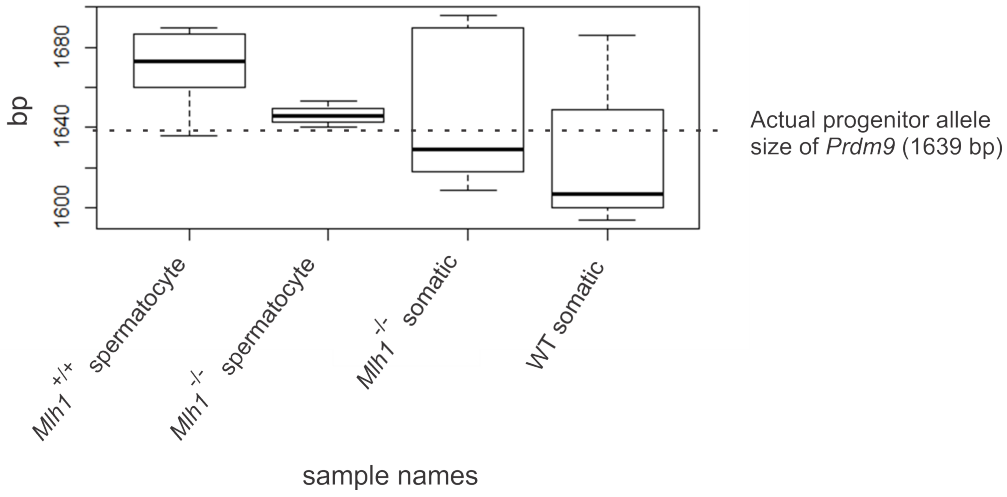

Figure 3.8: Boxplot of progenitor *Prdm9* allele size deviation

For *Prdm9* locus, incase of increase in one repeat unit (i.e. 84 bp), the size of the amplicon will be 1720 bp and incase of a decrease 1552 bp.
Germline instability was observed in both $Mlh1^{-/-}$ spermatocyte and $Mlh1^{+/+}$ spermatocyte samples. A single mutant was observed in $Mlh1^{-/-}$ spermatocyte samples. The mutant allele is shown in figure 3.7(a). Size of the mutant

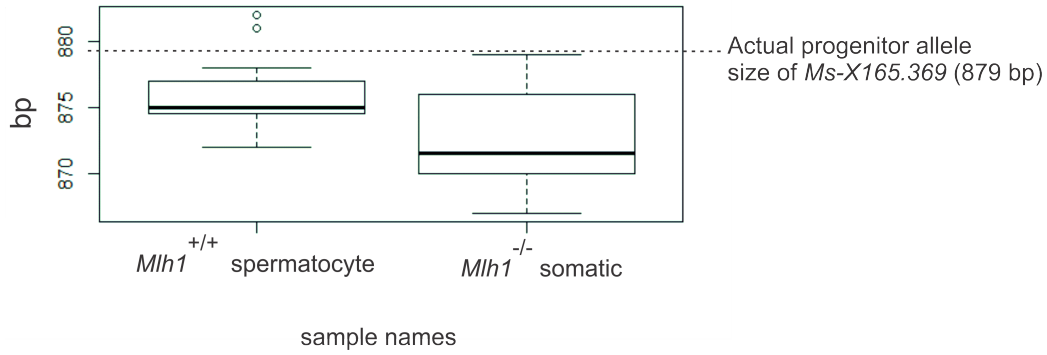

Figure 3.9: Boxplot of progenitor *Ms-X165.369* allele size deviation

was 1370 bp, which signifies loss of three repeat units. According to Poisson approximation, six molecules were analyzed. Therefore, germline mutation rate in *Mlh1* knockout mouse is 16.6% per gamete (1 out of 6). For $Mlh1^{+/+}$ spermatocyte samples, also a single mutant was observed. The mutant allele is shown in figure 3.7(b). This mutant had a size of 1898 bp. This signifies gain of three repeat units. For the blots for which we calculated the Poisson approximation, and for others considering we analyzed five molecules per PCR reaction, in total 72 molecules were analyzed. Therefore, mutation rate in $Mlh1^{+/+}$ mouse is 1.4% per gamete (1 out of 72). For the somatic samples i.e. $Mlh1^{-/-}$ somatic and WT somatic, no size shift was observed.
For *Ms-X165.369*, no instability was seen for both somatic and germline samples for any the genotype analyzed.

3.6 DNA binding motif prediction for the mutant PRDM9 zinc fingers and prediction of binding affinity of *Prdm9* alleles to the progenitor *Prdm9* allele's DNA binding site and to minisatellite *Ms-X165.369*

This analysis was performed using web based tool: A DNA binding site predictor for Cys2His2 Zinc Finger Proteins(http://zf.princeton.edu/index.php) (Persikov et al., 2009; Persikov and Singh, 2014).

The DNA binding site of a progenitor PRDM9 ZF is predicted to be GAACCC CCCCACCCCGCAGCAGCTGCTCGACATC as shown in the sequence logo (predicted DNA binding site) 3.10(a). The progenitor PRDM9 ZF has 11 zinc finger domain, each corresponding to a repeat unit. For the mutant alleles observed in this study, deletion or addition of repeat unit could be anywhere inside the repeat array. From Southern blot experiments, only the size shift of three repeat units was observed in the germline samples, but where exactly in the repeat array these increase or decrease occurs is not known. Here, I have shown three hypothetical cases of deletion and addition each, namely: decrease/increase of three repeat units from 3' end(Figure 3.10(b) and (e)), decrease/increase of three repeat units from 5' end (Figure 3.10(c) and (f)) and decrease/increase of three repeat units (5th,6th and 7th) from the middle (Figure 3.10(d) and (g))along with their sequence logos.
The prediction of binding affinity of *Prdm9* alleles to the progenitor allele's DNA binding site could only be made for mutant allele with deletion, as no scores were computed for the ones with additions. This is due to mathematical constraint in the algorithm which is discussed in section 4.4. All the *Prdm9* alleles with deletion were predicted to bind with the progenitor allele's DNA binding site, which is given as SVM score in Table 3.8. The SVM score is proportional to the DNA-protein binding affinity (Persikov et al., 2009). Similar prediction was made for binding affinity of *Prdm9* alleles to minisatellite *Ms-X165.369*. The scores are shown in Table 3.9.

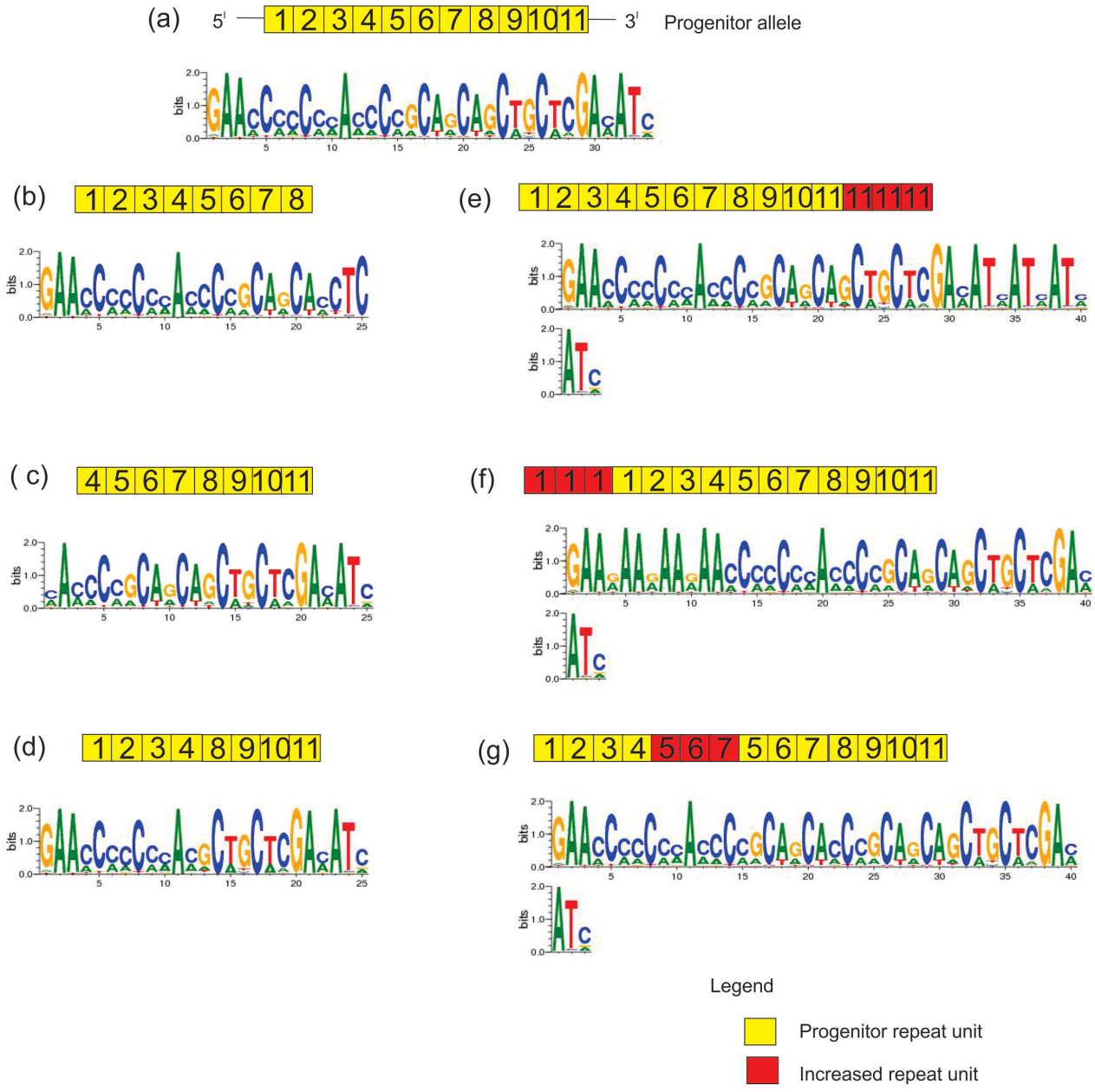

Figure 3.10: Sequence logo for progenitor PRDM9 zinc finger and six hypothetical cases of its mutant counterpart

***Prdm9* allele**	**SVM score**
progenitor	24.55
decrease in three repeat units in 5'	16.05
decrease in three repeat units in 3'	22.71
decrease in three repeat units from in between the array	9.92

Table 3.8: Predicted DNA-protein binding affinity of *Prdm9* alleles to the progenitor alleles DNA binding site

Prdm9 allele	**SVM score**
progenitor	7.87
decrease in three repeat units in 5'	4.82
decrease in three repeat units in 3'	10.94
decrease in three repeat units from in between the array	8.46
increase in three repeat units in 5'	8.18
increase in three repeat units in 3'	3.03
increase in three repeat units from in between the array	3.76

Table 3.9: Predicted DNA-protein binding affinity of *Prdm9* alleles to mini-isatellite *Ms-X165.369*

Chapter 4

Discussion

4.1 Problem due to low concentration of DNA samples

A major hurdle in this project was accurate quantification of DNA concentrations due to extremely low level of DNA samples. At extremely low level of DNA, there is no accurate way of quantifying DNA concentration, both nanodrop and Qubit were erroneous and measuring concentration by running in the gel was also only visible for concentration more than 1 ng. As shown in the microscopy images (Figure 3.2), the number of cell recovered from the filter for both the germline samples were extremely low, so obtaining a low yield of DNA was evident. More DNA could have been extracted from the entire testicular cell population, but we wanted a uniform cell population of spermatocytes from both the germline samples for our analysis. As such all the sperm cells from $Mlh1^{+/+}$ testicular cell population had to be separated as $Mlh1^{-/-}$ mouse do not produce any sperm, so that the comparison was more accurate. We were successful in separating most of the sperm to get an enriched spermatocyte cell population.

This problem of low level of DNA concentration could have been solved by starting with more starting material(i.e. more testes). But, we only possessed one of each $Mlh1^{-/-}$ and $Mlh1^{+/+}$ mouse, and producing more mutant mice would have taken more than four months.
Because of the same problem of low DNA concentrations, though our plan was to perform SP-PCR, for many of the PCR reactions we inadvertently ended up doing single molecule PCR (SM-PCR). Many groups have used SM-PCR in minisatellite instability studies. We, planned to use SP-PCR as we could screen more molecules per reaction. But, from our results it seem

only in presence of single molecule, instability is observed. This can either be coincidence or due to the reason mentioned in the introduction. The reason being that the predominant allele have dominance in amplification over the rare alleles in presence of high amount of predominant DNA molecule templates. Because of this, rare alleles will not amplify at all or if they amplify, it is not in the detectable range or dominant allele masks the rare allele during visualization.

4.2 Authenticity of size shift seen in lane 4 and 5 of blot 5

In blot number 5, size shift has been seen in Lane 3,4,5 on $Mlh1^{+/+}$ samples (Figure 3.7(b)). In careful evaluation of these three lanes, we can observe a gradual decrease in intensity of the fragments, which are of same size. This raises possibility that lane 4 and 5 (marked by degree symbol) are not genuine. There might have been an error in pipetting sample into gel; in doing so, small amount of PCR sample may have flooded into the neighboring lanes, or error might have occurred in the hybridization step. Whatever may be the reason, only size shift in lane 3 (marked by an asterisk) has been considered as actual increase in repeat units and calculation for instability has been done accordingly.

4.3 Evolutionary importance of the size shift in germline

Only the germline cell influences speciation. If any of the observed mutant spermatocytes develop into a sperm, and this sperm if and when generates offspring, this offspring will have only the mutant allele of respective minisatellite. Especially, in case of PRDM9, which is responsible for finding hotspots for recombination, either PRDM9 will bind to a totally new loci, thus totally new variation in chromosome occurs or there might be alteration in binding affinity of the protein or more severely total loss of bind affinity. In any case, it could either lead to totally new variation, thus in many generations leading to speciation or on the other hand, extinction of the species as a whole due to sterility caused by lack of HR.

4.4 Interpreting SVM scores

Score of DNA-protein binding affinity for progenitor PRDM9 to its own predicted DNA binding motif was 24.55. This is the maximum prediction score for this particular DNA-protein binding. If we consider this score to be 100%. Compared to this predicted result, as no DNA binding motif for PRDM9 in mouse has been defined till date. All the five minisatellites that were discovered had pretty low score, thus low binding affinity. The one scoring the highest i.e. *Ms-X165.369* had 67.9% lower binding affinity than the maximum predicted score. This signifies that though PRDM9 was predicted to bind to these minisatellites, it may or may not hold true in real biological context. A wet lab validation experiment must be done to come to any concrete conclusions.

For prediction of binding affinity of *Prdm9* alleles to the progenitor alleles' DNA binding site. The program could not predict DNA binding affinity for the mutant alleles with increase in the repeat units (Figure 3.10(e),(f),(g)). The affinity of all the mutant protein with deletion were less than that of the wild type PRDM9. Among the hypothetical cases shown in Figure 3.10, for the ones with deletions, the mutant PRDM9 with deletion in between the repeat array was predicted to have least binding affinity with 60% decrease in binding affinity and one with deletion at 3' end had the highest binding affinity with only 10% less affinity than wild type PRDM9. These results show that decrease in the repeat units form in between is more severe than others when it comes to binding affinity of the protein to its DNA motif. The program could not predict DNA binding affinity for the mutant alleles with increase in the repeat units. The calculations for SVM score is performed assuming each zinc finger domain binds to four nucleotides of DNA motif, when a zinc finger with three additional domain is used as an input along with the DNA motif of the wild type PRDM9, 12 base pairs in the DNA motif is less as such the inputs are not accepted.

In prediction of binding affinity of mutant *Prdm9* alleles to the minisatellite *Ms-X165.369*, compared to progenitor allele, increase in DNA-protein binding affinity was predicted in three cases: decrease in three repeat units at 3', decrease in three repeat units from in between the array and increase in three repeat units in 5'. The greatest increase in DNA-protein binding affinity among the three was for *Prdm9* allele with decrease in three repeat units at 3' which was predicted to be 39% higher. And for those alleles for which decrease in the affinity was predicted, the greatest decrease was predicted for

allele with increase in three repeat units in 3' with 51.5% decrease.

4.5 Comparison of our result with the published minisatellite instability in wild type

As mentioned in the introduction the most unstable locus known to date in wild type mouse is minisatellite locus Ms6-hm having germline mutation rate of 2.5% per gamete (Kelly et al., 1989) and one of the known stable mouse minisatellite is Mms80 with germline mutation rate at or below 5×10^{-6} per sperm (Bois et al., 2002). From our study, for *Prdm9* locus, we found the rate of germline mutation in $Mlh1^{+/+}$ mouse to be 1.4% per gamete. This is the wild type genotype. As such, rate of germline mutation at *Prdm9* locus for wild type mouse was found to be high, but not as high as the known most unstable locus. Also, in our study, no somatic mutation was observed in wild type mouse at *Prdm9* locus. This was a contrasting result to the most unstable locus, as somatic minisatellite instability of 2.8 % has been reported for *Ms6hm* locus (Kelly et al., 1989).
For minisatellite *Ms-X165.369*, no germline mutation was observed in the wild type mouse. This result shows that this locus is more stable in germline cells of wild type mouse than the known stable.

4.6 Comparison of our result with the published microsatellite instability in $Mlh1^{-/-}$ mice

In the study by Baker et. al. in characterized microsatellite in spermatocytes of $Mlh1^{-/-}$ mice , they found mutation in 14% of microsatellite locus *D9Mit67* and 18.3% of microsatellite locus *D1Mit355* (Baker et al., 1996). In our study for the same mice genotype we found germline minisatellite instability at locus *Prdm9* to be 16.6 % per gamete. These results show that germline mutation is quite similar for the two microsatellite and the minisatellite locus *Prdm9*. The somatic mutation for the same genotype for these two microsatellites were higher than the germline, 19% and 20.3% for *D9Mit67* and *D1Mit355* respectively. This was a contrasting result to what we observed for *Prdm9*. For *Prdm9* locus we did not observe any somatic mutation for $Mlh1^{-/-}$ mice. This suggests that *Prdm9* minisatellite locus is more stable in somatic cells than the two microsatellites.

For minisatellite *Ms-X165.369*, as no germline or somatic mutation was observed in the $Mlh1^{-/-}$ mouse. This minisatellite is more stable than the above mentioned microsatellites, in both germline and somatic tissues.

4.7 Comparison of minisatellite instability at *Prdm9* with *Ms-X165.369*

Instability was observed in both the genotypes at *Prdm9* locus in the germline samples, while no instability was observed at *Ms-X165.369* for the same genotype and samples. No somatic instability was observed for both the minisatellites in any of the genotype. This results shows instability is higher at *Prdm9* locus in the germline cells, and both are equally stable in the somatic tissues. This comparison is particularly interesting because *Ms-X165.369* is a potential DNA binding site for PRDM9 in germline cells and both are minisatellites.
The DNA-protein binding affinity could be altered either by instability in the DNA binding domain or in the DNA binding site. As germline Instability at minisatellite locus *Prdm9* was seen to be higher. We can come to a conclusion that change in DNA-protein binding affinity is more probable due to mutation in minisatellite *Prdm9* than in *Ms-X165.369*. Specially, among PRDM9 mutants, mutant with decrease in three repeat units at 3' end and increase in three repeat units in 3' end were more interesting as increase and decrease of DNA-protein binding affinity by 39% and 51.5% respective was predicted at these locus respectively, compared to the progenitor allele.

4.8 Germline and somatic minisatellite instability in $Mlh1^{-/-}$ and $Mlh1^{+/+}$ mice

Very little is known about minisatellite instability in absence of functional MLH1 protein. As mentioned in section 1.9, is it known that minisatellite instability is observed in absence of functional MLH1, but exact rate of instability is not known. Also, there has not been any study focusing on germline minisatellite instability in presence and absence of MLH1 protein in mouse. This study was first of its kind. We found germline mutation rates at *Prdm9* locus in *Mlh1* knockout mouse to be 16.6% per gamete and for $Mlh1^{+/+}$ mouse to be 1.4% per gamete. Somatic instability was not observed for both the genotype at this locus. This results suggest that functional MLH1 is required to maintain germline stability at *Prdm9* locus but is not essential

for its somatic stability. As, *Prdm9* gene is only expressed in germline cells (Segurel et al., 2011), stability of this gene in somatic tissues is not as much essential than in the germline. Also, in one genotype, instability caused loss of repeats while in other gain of repeats was observed. This suggest that the underlying mechanisms of instability for the two genotype may be different. Instability was not observed in both somatic and germline samples of any genotype at minisatellite locus *Ms-X165.369*.

Chapter 5

Limitations of this project and Future works

It would have been really interesting to assay germline minisatellite instability at the known stable (*Mms80*) and most unstable (*Ms6hm*) loci in absence of functional MLH1. But, due to limited-time frame of the project we could not perform the analysis. And for those minisatellite in which we analyzed the germline minisatellite instability, the sample size was too low.

We were only able to successfully analyzed total of 6 and 72 molecules for $Mlh1^{-/-}$ and $Mlh1^{+/+}$ spermatocyte samples respectively for *Prdm9* locus , in which we observed two mutants. Only a total of 81 PCR out of 151 PCR performed gave a positive PCR reaction for this locus. And for *Ms-X165.369* only 12 PCR for each sample was performed, all were positive PCR. These numbers are very small compared to other similar studies. Most of the minisatellite instability studies using either SM-PCR or SP-PCR have performed at least 135 to 400 positive PCR reactions per sample.

This thesis work was more of optimization of the protocol. As now the pipeline for analysis has been setup, the future work would be to analyze more samples. The sample size must be scaled up to get more accurate and more reliable results, also the problem of low concentration of DNA samples must also be taken care of. Prdm9 consist of SNPs (Appendix Figure A.4). If the repeat units that has been lost are the ones with SNPs or the canonical one is not known and is beyond this project's scope. But, this question can be answered in future using MVR-PCR.

Bibliography

L. A. Aaltonen, P. Peltomaki, F. S. Leach, P. Sistonen, L. Pylkkanen, J. P. Mecklin, H. Jarvinen, S. M. Powell, J. Jen, and S. R. Hamilton. Clues to the pathogenesis of familial colorectal cancer. *Science (New York, N.Y.)*, 260(5109):812–816, May 7 1993.

Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular biology of the cell*. Garland, 4 edition, 2002. ISBN 0815332181.

B. Alver, P. A. Jauert, L. Brosnan, M. O'Hehir, B. Vandersluis, C. L. Myers, and D. T. Kirkpatrick. A whole genome screen for minisatellite stability genes in stationary phase yeast cells. *G3 (Bethesda, Md.)*, Mar 11 2013.

J. W. Bacher, W. M. Abdel Megid, M. G. Kent-First, and R. B. Halberg. Use of mononucleotide repeat markers for detection of microsatellite instability in mouse tumors. *Molecular carcinogenesis*, 44(4):285–292, Dec 2005.

S. M. Baker, A. W. Plug, T. A. Prolla, C. E. Bronner, A. C. Harris, X. Yao, D. M. Christie, C. Monell, N. Arnheim, A. Bradley, T. Ashley, and R. M. Liskay. Involvement of mouse mlh1 in dna mismatch repair and meiotic crossing over. *Nature genetics*, 13(3):336–342, Jul 1996.

F. Baudat, J. Buard, C. Grey, A. Fledel-Alon, C. Ober, M. Przeworski, G. Coop, and B. de Massy. Prdm9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science (New York, N.Y.)*, 327(5967):836–840, Feb 12 2010.

G. Benson. Tandem repeats finder: a program to analyze dna sequences. *Nucleic acids research*, 27(2):573–580, Jan 15 1999.

I. L. Berg, R. Neumann, K. W. Lam, S. Sarbajna, L. Odenthal-Hesse, C. A. May, and A. J. Jeffreys. Prdm9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature genetics*, 42 (10):859–863, Oct 2010.

M. L. Bochman, K. Paeschke, and V. A. Zakian. Dna secondary structures: stability and function of g-quadruplex structures. *Nature reviews.Genetics*, 13(11):770–780, Nov 2012.

P. Bois, J. D. Stead, S. Bakshi, J. Williamson, R. Neumann, B. Moghadaszadeh, and A. J. Jeffreys. Isolation and characterization of mouse minisatellites. *Genomics*, 50(3):317–330, Jun 15 1998.

P. R. Bois, G. R. Grant, and A. J. Jeffreys. Minisatellites show rare and simple intra-allelic instability in the mouse germ line. *Genomics*, 80(1): 2–4, Jul 2002.

F. Bonhomme, E. Rivals, A. Orth, G. R. Grant, A. J. Jeffreys, and P. R. Bois. Species-wide distribution of highly polymorphic minisatellite markers suggests past and present genetic exchanges among house mouse subspecies. *Genome biology*, 8(5):R80, 2007.

J. Buard, E. Rivals, D. Dunoyer de Segonzac, C. Garres, P. Caminade, B. de Massy, and P. Boursot. Diversity of prdm9 zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite. *PloS one*, 9(1):e85021, Jan 13 2014.

M. G. Coleman, A. C. Gough, D. J. Bunyan, D. Braham, D. M. Eccles, and J. N. Primrose. Minisatellite instability is found in colorectal tumours with mismatch repair deficiency. *British journal of cancer*, 85(10):1486–1491, Nov 16 2001.

Y. E. Dubrova, M. Plumb, J. Brown, and A. J. Jeffreys. Radiation-induced germline instability at minisatellite loci. *International journal of radiation biology*, 74(6):689–696, Dec 1998.

S. Eaker, J. Cobb, A. Pyle, and M. A. Handel. Meiotic prophase abnormalities and metaphase cell death in mlh1-deficient mouse spermatocytes: insights into regulation of spermatogenic progress. *Developmental biology*, 249(1):85–95, Sep 1 2002.

W. Edelmann, P. E. Cohen, M. Kane, K. Lau, B. Morrow, S. Bennett, A. Umar, T. Kunkel, G. Cattoretti, R. Chaganti, J. W. Pollard, R. D. Kolodner, and R. Kucherlapati. Meiotic pachytene arrest in mlh1-deficient mice. *Cell*, 85(7):1125–1134, Jun 28 1996.

W. Edelmann, K. Yang, M. Kuraguchi, J. Heyer, M. Lia, B. Kneitz, K. Fan, A. M. Brown, M. Lipkin, and R. Kucherlapati. Tumorigenesis in mlh1 and

mlh1/apc1638n mutant mice. *Cancer research*, 59(6):1301–1307, Mar 15 1999.

C. D. Glen, A. G. Smith, and Y. E. Dubrova. Single-molecule pcr analysis of germ line mutation induction by anticancer drugs in mice. *Cancer research*, 68(10):3630–3636, May 15 2008.

D. B. Goldstein. Islands of linkage disequilibrium. *Nature genetics*, 29(2): 109–111, Oct 2001.

D. H. Gutmann, E. Winkeler, O. Kabbarah, N. Hedrick, S. Dudley, P. J. Goodfellow, and R. M. Liskay. Mlh1 deficiency accelerates myeloid leukemogenesis in neurofibromatosis 1 (nf1) heterozygous mice. *Oncogene*, 22 (29):4581–4585, Jul 17 2003.

D. D. Heath, G. K. Iwama, and R. H. Devlin. Pcr primed with vntr core sequences yields species specific patterns and hypervariable probes. *Nucleic acids research*, 21(24):5782–5785, Dec 11 1993.

RexA. Hess and LuizRenato de Franca. Spermatogenesis and cycle of the seminiferous epithelium. In C.Yan Cheng, editor, *Molecular Mechanisms in Spermatogenesis*, volume 636 of *Advances in Experimental Medicine and Biology*, pages 1–15. Springer New York, 2008. ISBN 978-0-387-79990-2. doi: 10.1007/978-0-387-09597-4_1. URL `http://dx.doi.org/10.1007/978-0-387-09597-4_1`.

A. Hochwagen and G. A. Marais. Meiosis: a prdm9 guide to the hotspots of recombination. *Current biology : CB*, 20(6):R271–4, Mar 23 2010.

B. Hopkins, N. J. Williams, M. B. Webb, P. G. Debenham, and A. J. Jeffreys. The use of minisatellite variant repeat-polymerase chain reaction (mvr-pcr) to determine the source of saliva on a used postage stamp. *Journal of forensic sciences*, 39(2):526–531, Mar 1994.

A. J. Jeffreys. Spontaneous and induced minisatellite instability in the human genome. *Clinical science (London, England : 1979)*, 93(5):383–390, Nov 1997.

A. J. Jeffreys, V. Wilson, S. L. Thein, D. J. Weatherall, and B. A. Ponder. Dna "fingerprints" and segregation analysis of multiple markers in human pedigrees. *American Journal of Human Genetics*, 39(1):11–24, Jul 1986.

A. J. Jeffreys, V. Wilson, R. Kelly, B. A. Taylor, and G. Bulfield. Mouse dna 'fingerprints': analysis of chromosome localization and germ-line stability

of hypervariable loci in recombinant inbred strains. *Nucleic acids research*, 15(7):2823–2836, Apr 10 1987.

A. J. Jeffreys, A. MacLeod, K. Tamaki, D. L. Neil, and D. G. Monckton. Minisatellite repeat coding as a digital approach to dna typing. *Nature*, 354(6350):204–209, Nov 21 1991.

A. J. Jeffreys, J. K. Holloway, L. Kauppi, C. A. May, R. Neumann, M. T. Slingsby, and A. J. Webb. Meiotic recombination hot spots and human dna diversity. *Philosophical transactions of the Royal Society of London.Series B, Biological sciences*, 359(1441):141–152, Jan 29 2004.

Alec J. Jeffreys and Rita Neumann. Somatic mutation processes at a muman minisatellite. *Human molecular genetics*, 6(1):129–136, January 01 1997.

Alec J. Jeffreys, Rita Neumann, and Victoria Wilson. Repeat unit sequence variation in minisatellites: A novel source of dna polymorphism for studying variation and mutation by single molecule analysis. *Cell*, 60(3):473–485, 2/9 1990.

C. Julier, B. de Gouyon, M. Georges, J. L. Guenet, Y. Nakamura, P. Avner, and G. M. Lathrop. Minisatellite linkage maps in the mouse by cross-hybridization with human probes containing tandem repeats. *Proceedings of the National Academy of Sciences of the United States of America*, 87 (12):4585–4589, Jun 1990.

S. H. Jun, T. G. Kim, and C. Ban. Dna mismatch repair system. classical and fresh roles. *The FEBS journal*, 273(8):1609–1619, Apr 2006.

L. Kauppi, A. J. Jeffreys, and S. Keeney. Where the crossovers are: recombination distributions in mammals. *Nature reviews.Genetics*, 5(6):413–424, Jun 2004.

R. Kelly, G. Bulfield, A. Collick, M. Gibbs, and A. J. Jeffreys. Characterization of a highly unstable mouse minisatellite locus: evidence for somatic mutation during early development. *Genomics*, 5(4):844–856, Nov 1989.

R. Kelly, M. Gibbs, A. Collick, and A. J. Jeffreys. Spontaneous mutation at the hypervariable mouse minisatellite locus ms6-hm: flanking dna sequence and analysis of germline and early somatic mutation events. *Proceedings.Biological sciences / The Royal Society*, 245(1314):235–245, Sep 23 1991.

O. Kikin, L. D'Antonio, and P. S. Bagga. Qgrs mapper: a web-based server for predicting g-quadruplexes in nucleotide sequences. *Nucleic acids research*, 34(Web Server issue):W676–82, Jul 1 2006.

P. Kumar, V. K. Yadav, A. Baral, P. Kumar, D. Saha, and S. Chowdhury. Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. *Nucleic acids research*, 39(18):8005–8016, Oct 2011.

M. A. Latif, M. Rafii Yusop, M. Motiur Rahman, and M. R. Bashar Talukdar. Microsatellite and minisatellite markers based dna fingerprinting and genetic diversity of blast and ufra resistant genotypes. *Comptes rendus biologies*, 334(4):282–289, Apr 2011.

A. Loukola, R. Salovaara, P. Kristo, A. L. Moisio, H. Kaariainen, H. Ahtola, M. Eskelinen, N. Harkonen, R. Julkunen, E. Kangas, S. Ojala, J. Tulikoura, E. Valkamo, H. Jarvinen, J. P. Mecklin, A. de la Chapelle, and L. A. Aaltonen. Microsatellite instability in adenomas as a marker for hereditary nonpolyposis colorectal cancer. *The American journal of pathology*, 155(6): 1849–1853, Dec 1999.

L. L. Mays-Hoopes, J. Bolen, A. D. Riggs, and J. Singer-Sam. Preparation of spermatogonia, spermatocytes, and round spermatids for analysis of gene expression using fluorescence-activated cell sorting. *Biology of reproduction*, 53(5):1003–1011, Nov 1995.

P. Modrich and R. Lahue. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annual Review of Biochemistry*, 65: 101–133, 1996.

D. G. Monckton, K. Tamaki, A. MacLeod, D. L. Neil, and A. J. Jeffreys. Allele-specific mvr-pcr analysis at minisatellite d1s8. *Human molecular genetics*, 2(5):513–519, May 1993.

M. E. Moynahan and M. Jasin. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature reviews.Molecular cell biology*, 11(3):196–207, Mar 2010.

U. G. Mueller and L. L. Wolfenbarger. Aflp genotyping and fingerprinting. *Trends in ecology evolution*, 14(10):389–394, Oct 1999.

H. Nybom, K. Weising, and B. Rotter. Dna fingerprinting in botany: past, present, future. *Investigative genetics*, 5(1):1–2223–5–1, Jan 3 2014.

P. L. Oliver, L. Goodstadt, J. J. Bayes, Z. Birtle, K. C. Roach, N. Phadnis, S. A. Beatson, G. Lunter, H. S. Malik, and C. P. Ponting. Accelerated evolution of the prdm9 speciation gene across diverse metazoan taxa. *PLoS genetics*, 5(12):e1000753, Dec 2009a.

P. L. Oliver, L. Goodstadt, J. J. Bayes, Z. Birtle, K. C. Roach, N. Phadnis, S. A. Beatson, G. Lunter, H. S. Malik, and C. P. Ponting. Accelerated evolution of the prdm9 speciation gene across diverse metazoan taxa. *PLoS genetics*, 5(12):e1000753, Dec 2009b.

E. D. Parvanov, P. M. Petkov, and K. Paigen. Prdm9 controls activation of mammalian recombination hotspots. *Science (New York, N.Y.)*, 327 (5967):835, Feb 12 2010.

A. V. Persikov and M. Singh. De novo prediction of dna-binding specificities for cys2his2 zinc finger proteins. *Nucleic acids research*, 42(1):97–108, Jan 2014.

A. V. Persikov, R. Osada, and M. Singh. Predicting dna recognition by cys2his2 zinc finger proteins. *Bioinformatics (Oxford, England)*, 25(1), Jan 1 2009.

Aris Polyzos, Craig Parfett, Caroline Healy, George Douglas, and Carole Yauk. A single-molecule pcr approach to the measurement of induced expanded simple tandem repeat instability in vitro. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 594(12): 93–100, 2/22 2006.

C. P. Ponting. What are the genomic drivers of the rapid evolution of prdm9? *Trends in genetics : TIG*, 27(5):165–171, May 2011.

T. A. Prolla, D. M. Christie, and R. M. Liskay. Dual requirement in yeast dna mismatch repair for mlh1 and pms1, two homologs of the bacterial mutl gene. *Molecular and cellular biology*, 14(1):407–415, Jan 1994.

G. F. Richard, A. Kerrest, and B. Dujon. Comparative genomics and molecular dynamics of dna repeats in eukaryotes. *Microbiology and molecular biology reviews : MMBR*, 72(4):686–727, Dec 2008.

Cheryl L. Scudamore. A practical guide to the histology of the mouse, 2014. ID: 854285818.

L. Segurel, E. M. Leffler, and M. Przeworski. The case of the fickle fingers: how the prdm9 zinc finger protein specifies meiotic recombination hotspots in humans. *PLoS biology*, 9(12):e1001211, Dec 2011.

N. Srivastava and M. J. Raman. Homologous recombination-mediated double-strand break repair in mouse testicular extracts and comparison with different germ cell stages. *Cell biochemistry and function*, 25(1):75–86, Jan-Feb 2007.

K. Usdin. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome research*, 18(7):1011–1019, Jul 2008.

M. N. Weitzmann, K. J. Woodford, and K. Usdin. The mouse ms6-hm hypervariable microsatellite forms a hairpin and two unusual tetraplexes. *The Journal of biological chemistry*, 273(46):30742–30749, Nov 13 1998.

C. Wyman, D. Ristic, and R. Kanaar. Homologous recombination-mediated double-strand break repair. *DNA repair*, 3(8-9):827–833, Aug-Sep 2004.

X. Yao, A. B. Buermeyer, L. Narayanan, D. Tran, S. M. Baker, T. A. Prolla, P. M. Glazer, R. M. Liskay, and N. Arnheim. Different mutator phenotypes in mlh1- versus pms2-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America*, 96(12):6850–6855, Jun 8 1999.

C. Yauk. Monitoring for induced heritable mutations in natural populations: application of minisatellite dna screening. *Mutation research*, 411(1):1–10, Aug 1998.

Carole L. Yauk, Yuri E. Dubrova, Gemma R. Grant, and Alec J. Jeffreys. A novel single molecule analysis of spontaneous and radiation-induced mutation at a mouse tandem repeat locus. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 500(12):147–156, 3/20 2002.

Appendix A

Appendix

Name	Sequence
Ms6-hm_F	AGAGTTTCTAGTTGCTGTGA
Ms6-hm_R	GAGAGTCAGTTCTAAGGC
MMS80_F	ACTGTGACCTCATGTTGCTAGG
MMS80_R	CAACGTCTCCAAGGTAACAGAG
PRDM9_F	GAGAATTTGCAATGGGGCTTT
PRDM9_R	ATATGGAATGGAATCATCGC
X165.369F	ACTCTGATGTACGAGCCAGG
X165.369R	GCCTGAGTCTAGGGGTTCTC
MS6-hmF_Probe	CACTGAGGAGTCCTGCTTCC
MS6-hmR_Probe	GCCATTGTAGGCGGTGGTAT
MMS80_F_probe	GCTTGTTGGCGATAAGGCTG
MMS80_R_probe	AGCAACGTCTCCAAGGTAACA
PRDM9_Int.F	TGTGGTTTTATTGCTGTTGGCT
PRDM9_Int.R	AATCCAAACCCAGGCAGAGG

Table A.1: Primers designed for the project

Conponent	Concentration of Stock	Volume(μl)	Concentration in PCR reaction
Tris-HCL ph 8.8	2M	167	45 mM
Ammonium Sulphate	1M	83	11 mM
Magnesium Chloride	1M	33.5	4.5 mM
2-mercaptoethanol	100 %	3.6	6.7 mM
EDTA pH 8.0	10 mM	3.4	4.4 mM
dATP	100 mM	75	1 mM
dCTP	100 mM	75	1 mM
dGTP	100 mM	75	1 mM
dTTP	100 mM	75	1 mM
BSA	10 mg/mL	85	113 ug/mL
Total Volume		676	

Table A.2: 11.1X buffer components

Stock	**Volume in reaction of 10μl (10μl)**
11.1 X Buffer	0.9
0.2M tris HCl (pH 7.5)	0.6
10μM Forward Primer	0.5
10μM Reverse Primer	0.5
0.5 U Kapa Taq.	0.6
2.5 U Cloned PFU	0.48
Water	5.42
template DNA	1
Total Volume	10

Table A.3: PCR master mix components

Temperature	**Duration**	**Cycles**
96 °C	100 sec.	X 1
96°C 49°C 70°C	20 sec. 30 sec. 1 min.	X 30
70 °C	7 min.	X 1
4 °C	∞	

Table A.4: PCR program in process of optimization for probe making of *Ms6hm* locus

sample name	**amplicon size**
$Mlh1^{-/-}$ spermatocyte	1646 ± 6.5
$Mlh1^{+/+}$ spermatocyte	1670 ± 17
$Mlh1^{-/-}$ somatic	1642 ± 32
WT somatic	1620 ± 28

Table A.5: Amplicon size measured for progenitor *Prdm9* locus

sample name	**amplicon size**
$Mlh1^{+/+}$ spermatocytes	876 ± 3
$Mlh1^{-/-}$ somatic	873 ± 4

Table A.6: Amplicon size measured for progenitor *Ms-X165.369* locus

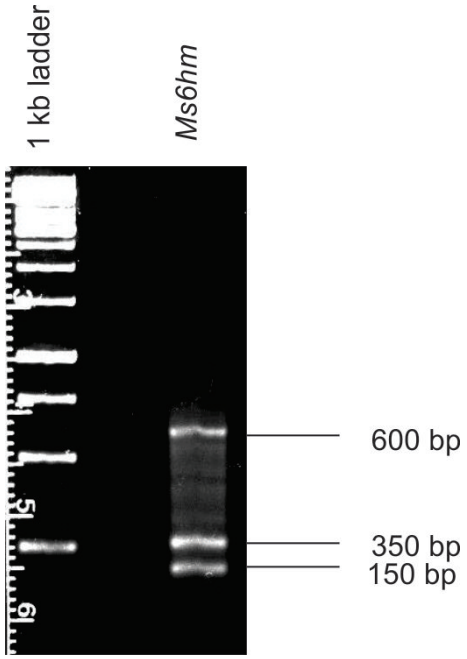

Figure A.1: Gel image of DNA fragments for *Ms6hm* after restriction digestion by AluI

Blot number	**sample name**	**a**	**b**	**P = - ln (a/b)**
1	$Mlh1^{-/-}$ spermatocyte	8	12	0.4
	$Mlh1^{+/+}$ spermatocyte	0	12	-
	$Mlh1^{-/-}$ somatic	0	12	-
2	WT somatic	13	15	0.14
	$Mlh1^{-/-}$ somatic	0	18	-
3	WT somatic	13	16	0.2
	$Mlh1^{-/-}$ somatic	3	17	1.7
4	$Mlh1^{+/+}$ spermatocyte	0	2	-
	WT somatic	1	13	2.56
5	$Mlh1^{+/+}$ spermatocyte	15	17	1.3
	$Mlh1^{-/-}$ spermatocyte	17	17	0

(a)

sample name	a	b	P = - ln(a/b)
$Mlh1^{-/-}$ spermatocyte	25	29	0.20
$Mlh1^{+/+}$ spermatocyte	15	17	0.13
$Mlh1^{-/-}$ somatic	3	17	1.73
WT somatic	27	44	0.50

(b)

Blot number	sample name	a	b	P = - ln(a/b)
6	$Mlh1^{-/-}$ spermatocytes	11	12	0.08
	$Mlh1^{+/+}$ spermatocytes	2	12	1.79
	$Mlh1^{-/-}$ somatic	0	12	-

(c)

Legend	
a	Number of negative PCR reactions
b	Total number of PCR
P	Poisson Approximation

Table A.7: Poisson approximation for number of amplifiable molecules for : (a)individual blots of *Prdm9* locus (b)All blots of *Prdm9* locus and (c)*Ms-X165.369* locus

locus name	Restriction enzymes	Cut size	coordinates in the locus
Prdm9	ApaLI	1208	793-2001
Mms80	MboII	1208	793-2001
Ms6hm	Alu1	497	1243-1740

Table A.8: Restriction enzymes for Minisatellites

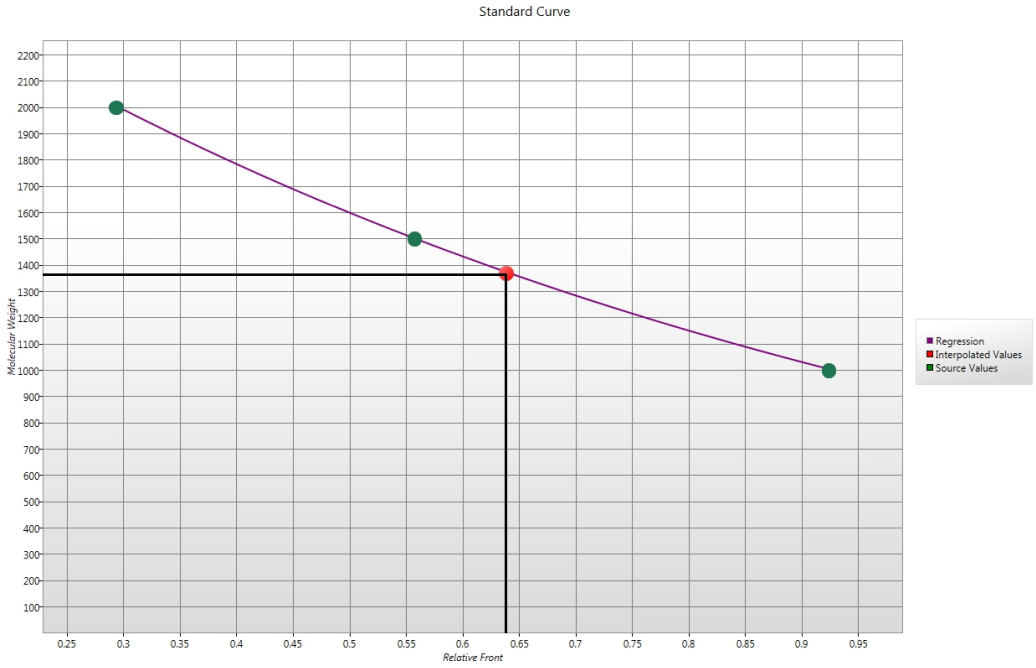

Figure A.2: Standard curve for measurement of migrating distance of mutant *Prdm9* allele of $Mlh1^{-/-}$ spermatocyte sample

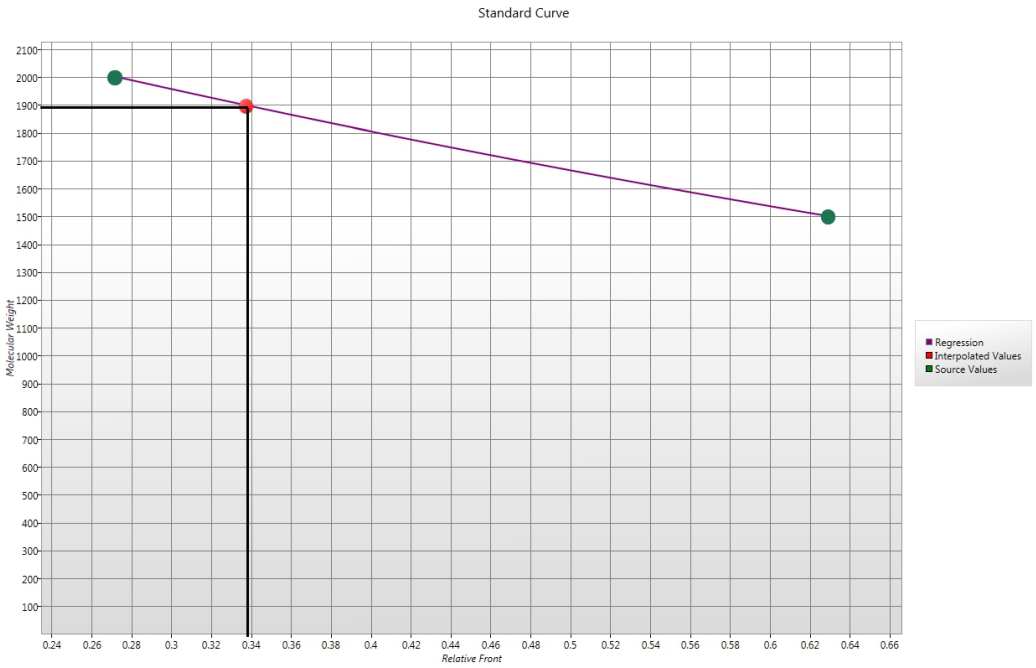

Figure A.3: Standard curve for measurement of migrating distance of mutant *Prdm9* allele of $Mlh1^{+/+}$ spermatocyte sample

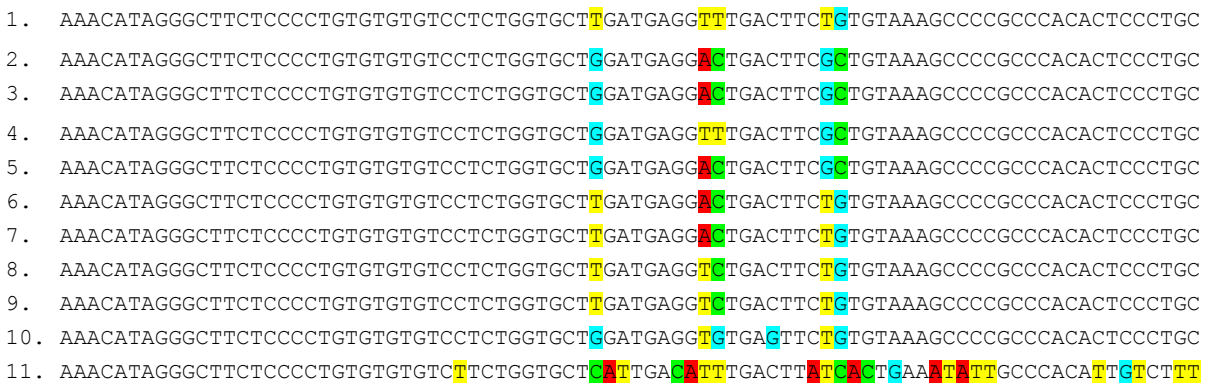

Figure A.4: 11 *Prdm9* repeat unit with SNPs

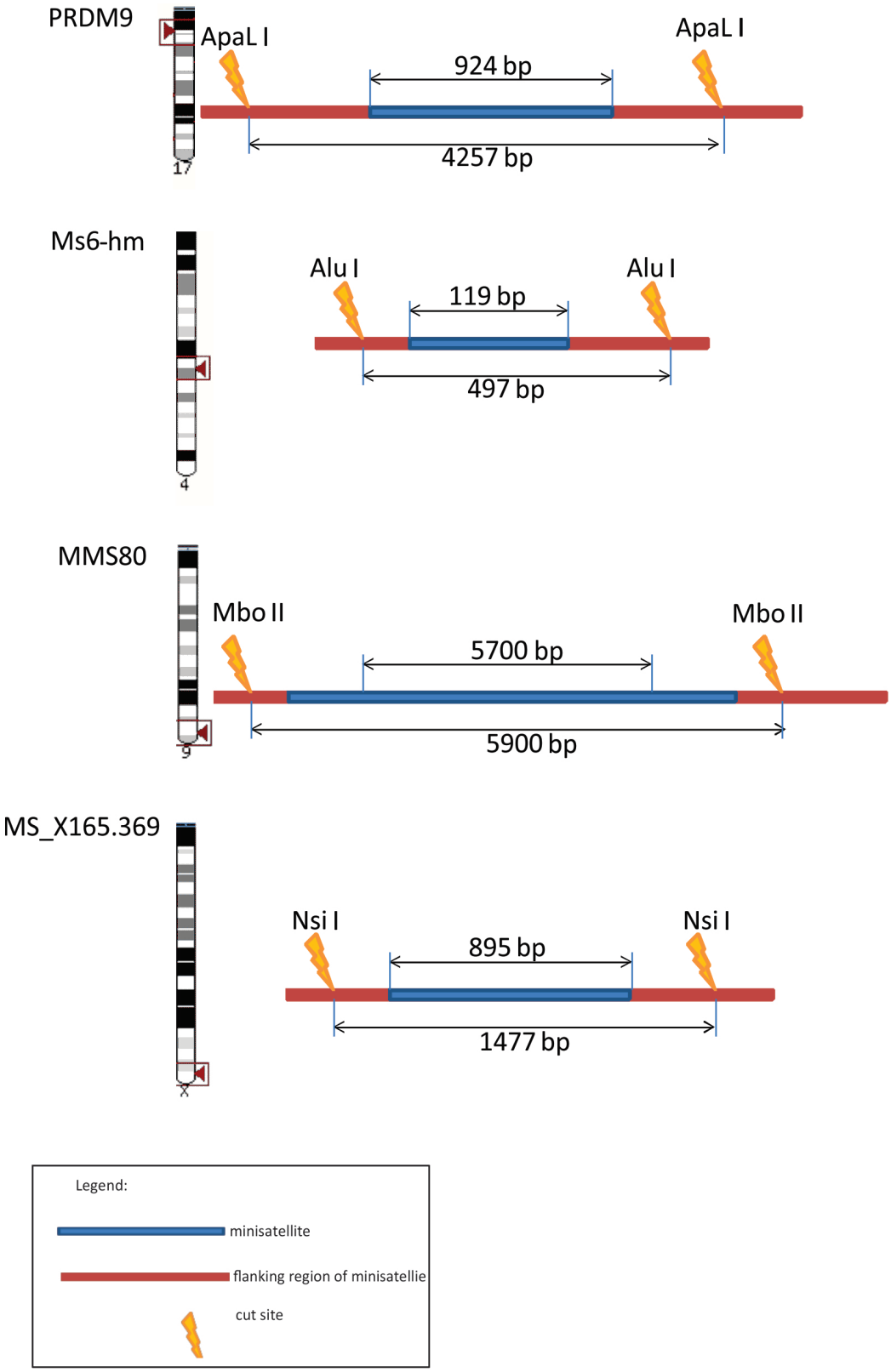

Figure A.5: Figurative representation of restriction enzymes for Minisatellites, along with location in the chromosome